



Уральский
федеральный
университет

имени первого Президента
России Б.Н. Ельцина

Институт социальных
и политических наук

Е. Л. МОГИЛЬЧАК

ВЫБОРОЧНЫЙ МЕТОД В ЭМПИРИЧЕСКОМ СОЦИОЛОГИЧЕСКОМ ИССЛЕДОВАНИИ

Учебное пособие

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б. Н. ЕЛЬЦИНА

Е. Л. Могильчак

ВЫБОРОЧНЫЙ МЕТОД В ЭМПИРИЧЕСКОМ СОЦИОЛОГИЧЕСКОМ ИССЛЕДОВАНИИ

Рекомендовано методическим советом УрФУ
в качестве учебного пособия для студентов, обучающихся
по программе бакалавриата по направлению подготовки
39.03.01 «Социология»

Екатеринбург
Издательство Уральского университета
2015

ББК С506я73-1
М742

Рецензенты:

кафедра теоретической и прикладной социологии
Уральского государственного педагогического университета
(заведующий кафедрой доктор социологических наук,
профессор Е. В. П р я м и к о в а);

Т. А. Л у г и н н а, кандидат социологических наук,
доцент кафедры социологии и социальной работы
Российского государственного
профессионально-педагогического университета

Научный редактор

А. В. М е р е н к о в, доктор философских наук, профессор

Могильчак, Е. Л.

М742 Выборочный метод в эмпирическом социологическом
исследовании : [учеб. пособие] / Е. Л. Могильчак ; М-во
образования и науки Рос. Федерации, Урал. федер. ун-т. –
Екатеринбург : Изд-во Урал. ун-та, 2015. – 120 с.

ISBN 978-5-7996-1479-9

В учебном пособии обобщен опыт преподавания дисциплины «Выборочный метод в социологическом исследовании» студентам-социологам. Дано понятие теории выборочного метода, показаны виды выборочных процедур, способы расчета случайных ошибок вероятностного отбора, систематические смещения и их источники. Учебное пособие снабжено большим количеством примеров, в том числе из практики проведения социологических исследований, приведены контрольные вопросы и практические задания.

Пособие предназначено для студентов социологических факультетов, факультетов политологии и социальной работы.

ББК С506я73-1

Предисловие

Проблематика, связанная с применением выборочного метода в социологическом исследовании, активно разрабатывалась в восьмидесятые годы прошлого века. Вышел целый ряд статистических учебников по выборочному методу. Также издавались монографии, которые были написаны социологами на примере конкретных социологических исследований, проводимых прежде всего Институтом социологии.

В настоящее время выборочному методу чаще всего бывают посвящены разделы в некоторых учебниках по социологии и статьи в социологических журналах, отдельные диссертационные работы. Нужно отметить, что определенная информация о применяемых в процессе общероссийских социологических исследований выборочных процедурах представляется на сайтах таких центров, как ФОМ, ВЦИОМ, «Левада-центр» и др. Но эта информация достаточно ограничена по своему объему. Указанной проблематикой также занимаются специалисты в области общей теории статистики, которыми издаются учебные пособия, предназначенные прежде всего для статистиков и экономистов.

Учебное пособие предназначено для студентов-социологов, обучающихся по направлению 39.03.01 «Социология» и изучающих учебный курс 27 МВ2.3. «Выборочный метод в социологическом исследовании».

Особенностью авторского подхода является четкое отделение выборочного метода от других способов проведения несплошных обследований. В учебном пособии описываются только такие процедуры, задачей которых является распространение полученных выводов на генеральную совокупность. Выборочный метод рассматривается как относящийся исключительно к *количественным методам* социологического исследования. Приступая к изучению курса, студенты должны освоить дисциплину «Прикладная статистика».

Книга написана в логике перехода от понятий теории выборочного метода к конкретным выборочным процедурам, которые применяются в эмпирическом социологическом исследовании. Предполагается, что учебное пособие создаст информационную основу для принятия студентом решения относительно выбора типа отбора в социологическом исследовании и реализации конкретных выборочных процедур.

Раздел 1

ВЫБОРОЧНОЕ ОБСЛЕДОВАНИЕ КАК РАЗНОВИДНОСТЬ НЕСПЛОШНОГО ОБСЛЕДОВАНИЯ. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЫБОРКИ

При проведении социологических исследований нередко возникает необходимость получения информации о больших совокупностях людей. Может показаться, что наиболее точным методом при этом является обследование всех людей, входящих в такие совокупности. Однако трудоемкость сплошного обследования чаще всего оказывается неприемлемой прежде всего из-за больших затрат времени и трудовых ресурсов. С увеличением численности обследуемых не только возрастают трудозатраты, но растет риск получения неточных данных, так как использование огромного числа анкетеров и кодировщиков снижает возможности тщательного контроля над их деятельностью.

Проведение сплошного обследования бывает возможным, если число обследуемых единиц не превышает нескольких тысяч, а сроки выполнения обследования жестко не лимитированы. Данный вид обследования может применяться и когда источником информации об индивидах являются документальные источники. Так, канадские исследователи Д. Бэрри и Р. Гоббинс в 1989 году исследовали политическую карьеру всех 3803 индивидов, которые служили в Канадской палате общин и сенате с 1867 по 1984 год¹. На основе анализа документов таким же образом могут быть исследованы все без исключения работники конкретного предприятия, организации. При небольшой численности трудового коллектива может быть применен метод опроса.

¹ См.: *Archer K., Gibbings R., Youngmen L. Explorations. A Navigators Guide to Quantative Research in Canadian Political Science.* N. Y., 1995. С. 209.

Обследование может быть не только сплошным, но и несплошным. Несплошное обследование предполагает изучение *части* единиц, входящих в интересующую социолога совокупность.

Несплошное обследование может иметь две разновидности:

1. Обследование, предполагающее распространение полученных результатов только на совокупность обследованных единиц².
2. Обследование, являющееся *выборочным*. Результаты данного обследования распространяются на более широкую совокупность, которая называется генеральной совокупностью.

Можно говорить о том, что данные два вида обследования отражают *два способа* формирования совокупности обследуемых единиц, то есть единиц, характеристики которых измеряются.

Охарактеризуем особенности *первой разновидности несплошного обследования*:

1. Применяемые методы формирования обследуемой совокупности относятся к *нестатистическим*.

2. Выборкой при проведении такого рода обследований называют *не* часть генеральной совокупности, результаты обследования которой распространяются на генеральную совокупность, а совокупность обследованных единиц. В случае применения метода опроса это совокупность всех опрошенных. Можно говорить о том, что расширительное толкование термина «выборка» противоречит классической теории статистики. Не предполагается выражение в числовой форме различий обследованной совокупности и совокупности, составляющей объект исследования. Это означает, что понятие «ошибка репрезентативности» не применяется.

Такого рода способы формирования обследуемой совокупности используются прежде всего в *качественных исследованиях*.

Кроме качественных исследований, способы с указанными выше свойствами используются в *целевом отборе*. Он предусмат-

² См.: Краткий обзор таких разновидностей отбора: *Девятко И. Ф.* Методы социологического исследования : учеб. пособие. М., 2002. С. 227–230 ; *Страус А., Корбин Дж.* Основы качественного исследования: обоснованная теория, процедуры и техники. М., 2001. С. 146–160 ; *Ковалев Е. М., Штейнберг И. Е.* Качественные методы в полевых социологических исследованиях. М., 1999. С. 71–78.

ривает отбор респондентов с заданными исследователем характеристиками в соответствии с целями исследования. К примеру, это могут быть посетители конкретного торгового центра в возрасте от 20 до 60 лет, имеющие высокий доход, постоянно проживающие в Москве³. Несмотря на то, что при осуществлении целевого отбора социолога интересует достаточно узкая категория населения, иногда она может иметь достаточно большую численность. Нередко целевая группа бывает труднодоступной, поэтому для формирования совокупности опрошенных в данном случае применяют специальные методы⁴.

Целевой отбор не предполагает существования процедуры, которая бы имела специфические свойства, характерные исключительно для данного вида отбора.

Следует отметить, что часто целевым отбором *неправомерно* называют любой способ формирования обследуемой совокупности, когда объектом исследования является достаточно узкая, специфическая часть населения. Однако если люди, относящиеся к данному объекту, отбираются по всем правилам формирования, например, общегородской выборки, то речь идет не о целевом отборе, а об обычной территориальной репрезентативной выборке. Просто генеральной совокупностью в данном случае будет не все взрослое население города, а, например, владельцы садовых и огородных участков, геймеры и т. д.

В ряде случаев целевым отбором пытаются назвать формирование обследуемой совокупности из определенного числа крупных единиц, которые выбираются исходя из предпочтений исследователя. Например, это наиболее крупные промышленные предприятия, на которых работает большая часть населения населенного пункта, либо несколько предприятий, которые являются *типичными* для данного города.

³ См.: Рогозин Д. М. Конформная выборка в торговых центрах // Социологический журнал. 2008. № 1. С. 22–47.

⁴ См.: Яковлева А. А. Исследования в труднодоступных группах: опыт использования выборки, управляемой респондентом, и выборки «место-время» // Социология 4М. 2011. № 33. С. 57–79.

Однако в данном случае происходит выделение не выборочной, а генеральной совокупности. Данную генеральную совокупность составляют работники, занятые *не* на всей совокупности предприятий города, а на отобранных социологом типичных (крупных) предприятиях, которые могут обследоваться в том числе на основе применения выборочного метода.

Вторая разновидность несплошного обследования основана на теории выборочного метода.

Выборочный метод – *количественный* метод отбора из совокупности единиц их части, результаты обследования которой распространяются на всю совокупность. В процессе формирования обследуемой совокупности в данном случае предполагается применение *статистических методов*. Чтобы делать достоверные выводы о целом по части, нужно обосновать критерии, по которым осуществляется отбор, и оценить точность такой экстраполяции.

Основные понятия теории выборочного метода

Генеральная совокупность – это совокупность единиц, относительно которой делаются выводы *выборочного обследования*. В качестве генеральной совокупности может выступать совокупность жителей страны, отдельного населенного пункта, совокупность работников предприятия и т. п.

Выборочная совокупность (выборка) – это часть генеральной совокупности, сформированная при помощи специальных методов и процедур, результаты обследования которой *распространяются на генеральную совокупность*.

Объем выборки – это число единиц, входящих в выборочную совокупность.

Объем генеральной совокупности – число единиц, входящих в генеральную совокупность.

Объем выборки или генеральной совокупности может выражаться не только числом людей, но и – в случае многоступенчатой выборки – числом избирательных участков, населенных пунктов,

то есть достаточно крупных единиц, которые состоят из единиц наблюдения.

Единицы наблюдения – элементы генеральной совокупности, свойства которых подвергаются непосредственному измерению. При опросах общественного мнения единицей наблюдения является отдельный респондент, при анализе документов единицей наблюдения может быть, например, отдельный экземпляр газеты.

В целях упрощения восприятия материала мы будем основываться на том положении, что единицей наблюдения является респондент, а выборочное исследование предполагает применение такого метода сбора информации, как опрос.

Единицы отбора – составные части (единицы) генеральной совокупности, последовательным отбором которых формируется выборочная совокупность. В качестве единиц отбора могут выступать как единицы наблюдения, так и более крупные единицы.

Одноступенчатая выборка – выборка, при которой выборочная совокупность формируется непосредственно из единиц наблюдения. Так, например, из списков работников завода сразу отбираются люди, принимающие участие в опросе.

Многоступенчатая выборка – выборка, при которой отбору единиц наблюдения предшествует отбор более крупных единиц, при этом на каждой ступени единица отбора меняется. Так, например, при *трехступенчатой* общегородской выборке на первой ступени могут отбираться избирательные участки, на второй ступени в каждом из отобранных избирательных участков отбираются квартиры, на третьей – в отобранных квартирах отбираются респонденты. На первой ступени единицей отбора является избирательный участок, на второй – квартира, на третьей – респондент.

Репрезентативность выборки – свойство выборочной совокупности воспроизводить характеристики генеральной совокупности.

Отклонение выборочных характеристик от генеральных называется **ошибкой репрезентативности** или ошибкой выборки. В случае количественных признаков, имеющих единицу измерения, ошибка выборки рассчитывается как разность генеральной и выборочной средней величины. В случае качественных признаков, то есть при-

знаков, не имеющих единицы измерения, ошибка выборки рассчитывается как *разность выборочной и генеральной доли* (процента).

Так, например, генеральная совокупность – совокупность работников завода. Ее численность – 5 тысяч человек. Из данной совокупности отобрано по таблице случайных чисел 450 человек. Следовательно, объем выборочной совокупности составил 450 человек, объем генеральной совокупности – 5000 человек. Опросив 450 человек, мы должны будем результаты опроса распространить на все 5 тысяч работников.

Так, если среди работников завода (генеральной совокупности) рабочие составляют *сорок два* процента, а в выборке их оказалось *тридцать восемь* процентов, то ошибка выборки по доле рабочих составит четыре процента ($42 - 38 = 4 \%$). Если выразить процентные показатели в виде долей, то можно представить расчет ошибки выборки следующим образом: $0,42 - 0,38 = 0,04$.

Источник получения информации о доле рабочих – отдел кадров предприятия. Данный источник предоставляет достоверные данные, отражающие состав совокупности, в которую входят абсолютно все работники, занятые на предприятии. Именно поэтому вероятность того, что *отклонение выборочной доли рабочих от их генеральной доли составляет именно 0,04, или 4 %*, является сто-процентной. Другими словами, мы полностью (на сто процентов), уверены в правильности данного статистического вывода.

Мы привели пример апостериорной (послеопытной) ошибки.

Апостериорная ошибка – это отклонение выборочного показателя от генерального, выраженное в виде *разности* выборочной и генеральной доли (средней), которое определяется путем сравнения генерального показателя, информация о котором получена из заслуживающих доверия источников, и выборочного показателя, полученного в результате опроса. (В социологии достаточно редко встречаются признаки, имеющие единицу измерения (количественные признаки), поэтому чаще всего ошибка выборки выражается в виде разности долей, а не средних.) В качестве указанных источников выступают органы статистики, другие государственные органы, отделы кадров предприятий.

Сравнив данные показатели (например, доли рабочих), мы на 100 % уверены, что значение ошибки составляет именно четыре процента.

Априорная ошибка – это отклонение выборочного показателя от генерального, выраженное в виде *разности* выборочной и генеральной доли и рассчитанное по формуле предельной ошибки случайной выборки. Данная ошибка относится исключительно к *вероятностному (случайному) отбору*. Вероятность правильности вывода о размере рассчитанной ошибки составляет *менее чем 100 %*, например, 95 % или 95 шансов из ста.

Случайная ошибка выборки – это ошибка вероятностного (случайного) отбора, которая представляет собой отклонение выборочного показателя от генерального, *выраженное в виде разности выборочной и генеральной доли (средней)* и вызванное исключительно несплошным характером обследования. Указанное отклонение *уменьшается с увеличением числа обследуемых единиц*.

Систематическая ошибка – отклонение выборочного показателя от генерального, выраженное в виде *разности* выборочной и генеральной доли (средней) и вызванное несоответствием способа формирования выборочной совокупности методическим правилам.

Отклонение выборочного показателя от генерального в данном случае *не* уменьшается с увеличением числа единиц отбора.

Так, проводя квартирный опрос, интервьюер опрашивает только тех членов семьи, которых застает дома, хотя он обязан отбирать их случайным образом из списка всех членов семьи. Это и приводит в данном случае к появлению систематической ошибки, которая *не* уменьшится с увеличением количества опрошенных.

Применение выборочного метода в процессе эмпирического социологического исследования имеет ряд этапов.

1. Создание проекта выборки.
2. Реализация проекта.
3. Оценка и коррекция апостериорных ошибок репрезентативности.

В *проекте выборки* определяется генеральная совокупность, описываются выборочные процедуры, рассчитываются объемы выборки.

Реализация проекта выборки органично входит в процесс сбора социологической информации и осуществляется путем выполнения анкетерами заданий, в которых указывается, где и каким образом следует осуществлять отбор респондентов.

Оценка и коррекция ошибок репрезентативности осуществляется после сбора информации. Определяются апостериорные ошибки выборки, а при необходимости осуществляется ремонт выборки.

Первое, что должен сделать исследователь в процессе составления проекта выборки, – определить генеральную совокупность. Это не всегда просто, так как зависит от целой совокупности факторов – предмета исследования, его задач, финансовых и организационных возможностей исследователя и т. п.

У ученого в процессе определения генеральной совокупности есть достаточно большая свобода выбора, главное требование при этом – четкость границ совокупности людей, на которую будут распространены выводы исследования и достаточная обеспеченность ресурсами – финансовыми и информационными – для осуществления такого проекта. В качестве генеральной совокупности может выступить все население в возрасте 18 лет и старше, проживающее на территории данного города. Если социолога интересует поведение или установки более узкой совокупности людей, то границы генеральной совокупности естественным образом сужаются. Например, это могут быть молодые люди в возрасте 18–29 лет, проживающие на территории данного населенного пункта. В качестве генеральной совокупности может выступать и население всей страны, работники конкретного предприятия и т. п.

Не нужно забывать, что в реальности в качестве генеральной совокупности в массовых опросах выступает не все население страны, области, города и т. п., а совокупность людей, из которой исключены некоторые специфические группы населения. Например, это срочные военнослужащие, тяжелобольные, а иногда жители труднодоступных поселений. В последнем случае об этом должна предоставляться информация исследовательской организацией.

Нужно особо отметить, что генеральная совокупность должна быть однородной. Это означает, что она не должна состоять из нескольких разнокачественных категорий людей, например, студентов дневных отделений вузов и работников промышленных предприятий данного города. В данном случае можно говорить о двух разных генеральных совокупностях.

Вопросы и задания для самоконтроля

1. Дайте определение понятия «выборочный метод».
2. Как соотносятся понятия «единица наблюдения» и «единица отбора»?
3. В чем отличие систематической ошибки выборки от случайной ошибки?
4. Перечислите этапы применения выборочного метода в процессе эмпирического социологического исследования.
5. Приведите примеры несплошных обследований, не являющихся выборочными.

Раздел 2

СЛУЧАЙНЫЕ И НЕСЛУЧАЙНЫЕ ВИДЫ ОТБОРА

Если генеральная совокупность определена, то необходимо решить, каким образом отбирать в выборку ее элементы, то есть описать выборочные процедуры.

Выборочная процедура – это последовательность действий при формировании выборочной совокупности.

Выборочные процедуры могут быть разделены на типы по двум основным критериям (кроме данных критериев, можно использовать другие средства классификации – например, разделить выборку на производственную и территориальную):

1) по роли вероятностных законов в процессе формирования выборки;

2) по количеству ступеней отбора.

Исходя из *первого критерия*, можно выделить:

- случайный (вероятностный) отбор (случайная выборка);
- неслучайный отбор.

Второй критерий предполагает выделение:

- одноступенчатой выборки;
- многоступенчатой выборки.

Типы отбора, выделенные по роли вероятностных законов в процессе формирования выборки

1. Случайные виды отбора

Слово «случайный» в повседневной лексике имеет значение, противоположное содержанию данного понятия в математике. Случайный (вероятностный) отбор должен осуществляться по жестким правилам, применение которых не допускает никакого произ-

вола и обеспечивает каждому элементу генеральной совокупности примерно равные с другими шансы попасть в выборочную совокупность. При неслучайном отборе эти шансы не равны друг другу.

Случайная выборка имеет следующие разновидности:

- простая случайная;
- систематическая (механическая);
- гнездовая (серийная, кластерная);
- стратифицированная (районированная или типическая).

Для построения случайной выборки любого вида требуется *основа выборки* – перечень всех элементов генеральной совокупности. Например, это может быть список всех работников завода. В качестве элементов, которые случайным образом отбираются, могут выступать как люди, так и более крупные единицы.

При больших объемах генеральной совокупности простая случайная выборка используется для отбора крупных единиц на первых ступенях многоступенчатой выборки. В любом случае для осуществления этого типа выборочной процедуры требуются таблицы случайных чисел.

Опишем виды вероятностного отбора, которые используются для отбора единиц наблюдения, а не более крупных единиц, то есть речь идет об *одноступенчатых выборках*.

Простая случайная выборка

Данный вид отбора осуществляется следующим образом:

1. Определяется объем выборки при помощи специальной формулы.
2. Составляется полный список элементов генеральной совокупности.
3. При помощи таблицы случайных чисел из списка отбираются рассчитанное ранее число людей, которые будут обследоваться.

Таблицы случайных чисел приводятся в специализированных математико-статистических изданиях и представляют собой ряды чисел, которые определены вероятностным способом. Использование для отбора таблиц с названием, отличным от вышеуказанного названия, исключается.

Фрагмент таблицы случайных чисел⁵

66194	28926	99547	16625	45515	67953	12108	78240	43195	24837
32511	70880	22070	52622	00833	88000	67299	68215	11274	55624
00833	88000	67299	68215	11274	55624	32991	17436	12111	86683
61270	58036	64192	90611	15145	01748	47189	99151	05755	03834

Числа в таблице указывают, под каким номером в списке следует включить в выборку человека. Если в списке 7900 человек, а отобрать нужно 500, то необходимо получить пятьсот четырехзначных чисел. Нужно особо отметить, что в том случае, если числовой ряд в таблице разбит на пятизначные числа, как в нашем случае, то мы не должны обращать внимание на разделительные вертикальные линии. Следует отбирать четырехзначные числа, одно за другим, если даже цифры окажутся в разных колонках. Так, например, после числа 6619 мы должны отобрать число 4289, несмотря на то, что первая цифра – 4 – стоит в первом столбце, а три последние цифры – 2, 8, 9 – оказались в следующей колонке.

Так как объем выборки – четырехзначное число, то номер каждого человека в списке тоже должен быть четырехзначным – от 0001 до 7900. Тогда в выборку попадут респонденты под номерами 6619, 4289, 2699, 5471, 6625 и т. д.

Некоторые числа *пропускаются*. Это происходит, если какое-то число встретилось второй раз либо среди потенциальных респондентов отсутствует человек под таким номером. Так, например, число 8240 пропускается, так как объем генеральной совокупности составляет 7900 человек, и потенциального респондента под номером 8240 просто не существует.

В результате выполнения указанных действий число отобранных номеров должно достигнуть требуемого объема выборки.

⁵ См.: Елисеева И. И., Юзбашев М. М. Общая теория статистики. М., 1996. С. 362.

Систематическая выборка

Данный вид отбора осуществляется следующим образом:

1. Рассчитывается объем выборки по специальной формуле.
2. Составляется полный список элементов генеральной совокупности.
3. Определяется шаг выборки.
4. По таблице случайных чисел определяется номер первого респондента.
5. Производится отбор из списка респондентов с данным шагом.

Список, из которого будет производиться отбор, должен быть алфавитным или другим «случайно» составленным, так как наличие в списке определенной последовательности, закономерности, может служить источником грубых систематических ошибок.

Шаг выборки рассчитывается делением объема генеральной совокупности на объем выборки (n). Если объем генеральной совокупности (N) составил 95 000, то нужно 95 000 разделить на 500.

$$\frac{N}{n} = \frac{95\,000}{500} = 190.$$

При делении мы получили 190, поэтому должны отбирать каждого стодевяностого человека. Поэтому если первый отобранный имеет номер 661 (первое в таблице случайных чисел число), то следующим будет респондент с номером 850 ($660 + 190 = 850$), затем – 1040-й в списке ($850 + 190 = 1040$) и т. д. Номер первого респондента, как и было сделано, следует определять по таблице случайных чисел.

Априорную ошибку систематической выборки рассчитывают по тем же формулам, что и в простой случайной выборке.

Следует отличать систематическую и простую случайную выборку как основные типы отбора в исследовании и как *вспомогательные*, входящие как компоненты в более сложные выборочные процедуры, например в многоступенчатую выборку на отдельных ступенях, а также гнездовую, стратифицированную выборку. В этом случае они могут применяться не только для отбора людей, но и квартир, населенных пунктов и т. д.

Так, систематический отбор квартир может быть осуществлен непосредственно анкетерами на заданных социологом сегментах территории. Например, может отбираться каждый десятый дом индивидуальной застройки на данном участке улицы.

Простая случайная выборка в качестве *вспомогательной процедуры* может применяться и в многоступенчатой выборке для отбора, например, избирательных участков.

В случае же применения систематической или простой случайной выборки как *основной процедуры* исследователем обязательно составляется полный перечень элементов генеральной совокупности.

Процедура, близкая к систематическому отбору, применяется при опросах избирателей при **выходе с избирательных участков** (exit poll). Избиратели отбираются визуально, без предварительного составления списков и опрашиваются интервьюером с интервалом 5–6 минут. В сельских населенных пунктах в первой половине дня интервал сокращается до 4–5 минут в связи с большим числом приходящих голосовать в данное время. Ясно, что, исходя из чисто формальных соображений, этот тип отбора нельзя назвать систематическим, так как интервьюер использует при отборе не установку «опрашивать каждого n -го избирателя», а «опрашивать избирателя один раз в n -е число минут». Однако при условии относительно равномерного потока избирателей в течение дня данный вид отбора можно приравнять к систематическому, который обеспечивает каждому избирателю, пришедшему голосовать, равные шансы попасть в выборку.

Гнездовая выборка

Гнездовая выборка применяется в том случае, если генеральная совокупность состоит из *небольших по численности* естественных групп элементов (серий, гнезд).

При использовании данного типа отбора необходимо:

1. Определить объем выборки (число гнезд).
2. Составить полный список гнезд, входящих в генеральную совокупность.

3. По таблице случайных чисел (или систематически) отобрать рассчитанное ранее число гнезд.

4. Выдать задание анкетерам на *сплошной* опрос всех людей, входящих в каждое из отобранных гнезд.

Чем большее число серий отбирается, тем ошибка гнездовой выборки меньше. Поэтому *гнезда должны быть небольшими по численности*, иначе осуществить обследование большого их числа будет невозможно. Гнездовую выборку следует применять в тех случаях, когда гнезда схожи по изучаемому признаку. Внутри же каждой из серий «разброс» значений признака должен быть максимальным.

Так, если исследуются ценностные ориентации школьников, то гнездовой отбор будет эффективен, если между классами, выступающими в качестве гнезд, не будет сильного различия по характеру данных ориентаций. Внутри же каждого класса должно быть представлено все разнообразие ценностных предпочтений.

Например, объектом исследования являются учащиеся девятих классов определенного города. Всего в школах города – 210 девятих классов. Объем выборки запланирован в количестве 30 классов. Разделим объем генеральной совокупности на объем выборочной совокупности. $210 : 30 = 7$. Поэтому из списка мы должны отобрать каждый седьмой класс, используя систематический отбор. То же самое можно сделать, используя таблицы случайных чисел. Получив список из 30 школьных классов, мы можем выдавать задание анкетерам на сплошной опрос в каждом классе, причем речь идет о *полном охвате всего списочного состава* школьников в каждой школе. Данный вид отбора может применяться только в случае практически стопроцентной посещаемости школьников.

Стратифицированная выборка

Стратифицированная выборка (районированная, расслоенная, иногда ее называют типической) – выборка, при которой, перед тем как начать отбор, генеральная совокупность разбивается на однородные части (страты) по определенному признаку, а затем из них производится систематический или простой случайный отбор.

Признак, по которому производится расслоение генеральной совокупности, называется *признаком расслоения*. Страты представляют собой группы единиц, имеющие разные значения признака расслоения. Если таким признаком является уровень образования, то формируется 4 типических группы (страты) – люди с образованием ниже среднего, со средним общим образованием, средним специальным и высшим образованием. Стратами могут быть и люди, проживающие в разных регионах или обучающиеся в разных вузах города. Тогда признаком расслоения будет региональная или вузовская принадлежность. Следует выбирать такие признаки расслоения, которые оказывают существенное влияние на *исследуемые признаки*. Исследуемыми являются признаки, измерение которых входит в задачи исследования, например, удовлетворенность населения различными аспектами жизни, электоральные предпочтения и т. п.

Лучше, если стратификация будет производиться по двумерному или трехмерному признаку, например, в двумерном случае – по полу и возрасту. Иногда данные признаки называют сопряженными. Тогда образуются различные половозрастные страты: мужчины 18–29 лет, женщины 18–29 лет, мужчины 30–49 лет, женщины 30–49 лет и т. д.

Если из каждой страты производится отбор пропорционально числу единиц в ней, то такой отбор называется пропорциональным стратифицированным.

В ряде случаев может производиться *непропорциональный* стратифицированный отбор. Он применяется, если нужно увеличить объем выборки из слишком маленькой страты или когда единственной целью исследования является сравнение страт. Непропорциональная стратификация может быть осуществлена и при так называемом **оптимальном размещении выборки**⁶, когда объем выборки из каждой страты увеличивается с увеличением степени неоднородности ее элементов. В данном случае необходимо знать дисперсию единиц, входящих в каждую страту. Чем выше дисперсия, тем выше объем выборки из страты.

⁶ Описание данного метода см., напр.: *Шварц Г.* Выборочный метод. Руководство по применению статистических методов оценивания / пер. с нем. М., 1978. С. 151–152.

Для осуществления *пропорциональной* стратифицированной выборки необходимо:

1. Определить общий объем выборки.
2. Выбрать признак расслоения.
3. Получить статистические данные о численности каждой страты в генеральной совокупности.
4. Рассчитать число людей, которое следует отобрать из каждой страты.
5. Составить полный список элементов (людей), входящих в каждую страту.
6. При помощи простой случайной или систематической выборки произвести отбор из каждой страты рассчитанного ранее числа элементов. Объем выборки из страт при данном виде отбора пропорционален числу людей, входящих в страту.

Например, генеральной совокупностью в исследовании является совокупность рядовых работников промышленного предприятия. Их число составляет 2000 человек. 1500 работников (75 %) являются рабочими, 400 (20 %) – инженерами и техниками, 100 (5 %) – служащими (табл. 2).

Т а б л и ц а 2

Распределение генеральной совокупности по 18 стратам

Должность работника	Общее количество работников	Цех № 1	Цех № 2	Отдел 1	Отдел 2	Отдел 3	Отдел 4
Рабочие	1500	760	580	4	4	–	–
Инженеры и техники	400	24	12	104	180	108	152
Служащие	100	16	8	12	16	12	8
Общий объем генеральной совокупности	2000	800	600	120	200	120	160

Объем выборки составляет 500 человек. Следовательно, из списков работников, полученных в отделе кадров предприятия, следует

при помощи таблицы случайных чисел отобрать 75 % рабочих ($500 \cdot 0,75 = 375$ человек), 20 % инженеров и техников ($500 \cdot 0,20 = 100$ человек) и 5 % служащих ($500 \cdot 0,05 = 25$ человек).

Если мы решили в целях увеличения уровня репрезентативности сделать признак расслоения двумерным, то проще, определив общий объем выборки, рассчитать число людей, которых нужно отобрать из каждой страты, следующим образом:

1. Определить долю, которую составляет общий объем выборки в объеме генеральной совокупности (n/N).
2. Умножить эту долю на численность каждой страты.

Разделим объем выборки на объем генеральной совокупности: $\frac{n}{N} = \frac{500}{2000} = 0,25$. Теперь можно число 0,25 умножать на каждую из внутриклеточных частот в таблице, представляющей генеральную совокупность.

Так, 760 мы умножим на 0,25 и получим 190. Значит, в цехе № 1 мы должны отобрать 190 рабочих. Таким же образом, произведя вычисления для каждой клетки таблицы, впишем рассчитанные числа в таблицу 3.

Структура выборочной совокупности будет в этом случае следующей:

Т а б л и ц а 3

Распределение выборочной совокупности по 18 стратам

Должность работника	Общее количество работников	Цех № 1	Цех № 2	Отдел 1	Отдел 2	Отдел 3	Отдел 4
Рабочие	375	190	145	1	1	–	–
Инженеры и техники	100	6	3	26	45	27	38
Служащие	25	4	2	3	4	3	2
Общий объем выборки	500	200	150	30	50	30	40

Процент, который составляют в объеме выборочной совокупности работники каждой страты, полностью совпадает с процентом данной страты в численности всех работников завода, то есть в генеральной совокупности.

Так число рабочих цеха № 1 в выборке – 190 человек, общий объем выборки – 500 человек, следовательно, процент рабочих данного цеха в выборочной совокупности составляет $\frac{190}{500} \cdot 100 = 38 \%$.

Этот процент идентичен соответствующему показателю по генеральной совокупности. $\frac{760}{2000} \cdot 100 = 38 \%$ (см. табл. 2).

То же самое можно сказать о каждой из оставшихся 17 страт, так, например, инженеры и техники отдела 4 составляют 7,6 % и в генеральной, и в выборочной совокупности, служащие отдела 3 – 0,8 % и т. д.

Теперь для цеха № 1 мы должны составить пофамильные списки 760 рабочих, 24 инженеров, 16 служащих и выбрать *по таблице случайных чисел* из первого списка 190 человек, из второго – 6 человек, из третьего – 4 человека. Точно также следует отобрать респондентов из списков рабочих, служащих и инженеров и техников второго цеха, а также каждого из четырех отделов.

В результате указанного отбора каждый человек, входящий в конкретную страту в генеральной совокупности, имеет *равные шансы с другими людьми, входящими в данную страту*, попасть в выборочную совокупность.

Объем стратифицированной выборки может быть меньше, чем при простом случайном отборе, так как уровень ее репрезентативности выше. Однако нужно учитывать, что выигрыш в репрезентативности достигается только в случае наличия тесной статистической связи между признаками расслоения и исследуемыми признаками. Если эта связь очень слабая, то выигрыш в репрезентативности по сравнению с простой случайной или систематической выборкой будет сведен к нулю. Так, например, если признаком расслоения является пол, а в результате опроса выяснилось, что пред-

почтения мужчин практически ничем не отличается от предпочтений женщин, то случайная ошибка данной выборки будет такой же, как при простом случайном отборе. Принимая во внимание данную возможность, лучше объем стратифицированной выборки принимать равным объему, требуемому при простой случайной выборке.

Все перечисленные нами разновидности случайной выборки, включая гнездовую и стратифицированную, являются **одноступенчатыми**. Так, отбор гнезд предполагает автоматическое включение в выборочную совокупность всех единиц наблюдения, входящих в эти гнезда. Также разделение генеральной совокупности на страты в процессе применения стратифицированной выборки представляет *не отдельную ступень отбора, а этап, предшествующий началу собственно формирования выборки* – число страт не уменьшается в процессе осуществления данной процедуры.

Нужно отметить, что при использовании **любой разновидности случайного отбора** обязательно должны использоваться повторные посещения отсутствующих респондентов. Обычно их число равно трем. Ясно, что при наличии возможности лучше делать большее число визитов. Однако именно после трех посещений их эффективность существенно снижается, и посещать респондента повторно становится нецелесообразным, в том числе и экономически.

Перечислим преимущества и ограничения описанных нами выборочных процедур.

1. Из всех перечисленных случайных выборочных процедур *стратифицированная выборка является наиболее затратной* – для ее формирования нужно иметь больше как финансовых, так и информационных ресурсов. Достигаемая *случайная ошибка* выборки в данном случае минимальная.

2. Гнездовая выборка является наименее затратной, но дает максимальную случайную ошибку (обоснование более высокой случайной ошибки гнездовой выборки приведено в разделе 4).

3. Недостатком как стратифицированной, так и простой случайной и систематической выборки является необходимость создания организационных условий для решения проблемы анонимности в процессе опроса.

4. Так как ошибка случайной выборки включает не только случайную, но и систематическую ошибку, то совокупная ошибка будет зависеть от возможностей реализации требований случайного отбора применительно к конкретному объекту исследования.

2. Неслучайные виды отбора

Как уже было отмечено выше, шансы попадания в выборку элементов генеральной совокупности при неслучайном отборе не равны друг другу и определяются специфическими характеристиками людей, позволяющими одним иметь большую вероятность попасть в выборку по сравнению с другими. Единственным видом неслучайного отбора, имеющим определенное математическое обоснование, является квотная выборка.

Квотная выборка

Квотная выборка формируется из единиц определенных категорий (квотных групп), которые должны быть представлены в пропорциях, *соответствующих пропорциям в генеральной совокупности*. Таким образом происходит целенаправленное согласование распределения определенных признаков в генеральной и выборочной совокупности. Эти признаки называются *квотными*. Они должны оказывать сильное влияние на исследуемые признаки, то есть признаки, измерение которых входит в задачи исследования. В качестве исследуемых признаков могут выступать мнения, оценки и т. п. В качестве квотных признаков чаще всего используются пол, возраст, образование, так как информация об их распределении в генеральной совокупности наиболее доступна. Квотные группы – это группы людей, которые имеют разные значения квотного признака. В данном случае *квотными группами* будут выступать мужчины и женщины, возрастные и образовательные группы.

В целях повышения репрезентативности рекомендуется формировать группы по сопряженным квотным признакам. Это означает, что каждая группа должна обладать сочетанием значений нескольких признаков – например, формируются группы женщин с высшим образованием в возрасте 40–49 лет, мужчин со средним специальным образованием в возрасте 20–29 лет и т. д. Данные

о количественном составе данных групп берутся в органах государственной статистики или отделе кадров предприятия. Цель выборочной процедуры заключается в том, чтобы обеспечить в выборке такое же процентное соотношение квотных групп, что и в генеральной совокупности. В данном случае квотные группы имеют разные значения *трехмерного* квотного признака – поло-образовательно-возрастной принадлежности.

Увеличение числа одномерных признаков, образующих многомерную квотную переменную, может уменьшить ошибку репрезентативности. С другой стороны, увеличиваются трудности «заполнения» множества квотных групп, некоторые из них к тому же могут оказаться малочисленными.

Необходимо особо отметить, что не следует выбирать квотные признаки, значения которых быстро меняются с течением времени, так как существует опасность использования в процессе отбора устаревших данных.

Для осуществления квотной выборки следует:

1. Определить общий объем выборки.
2. Выбрать квотные признаки.
3. Получить статистические данные о численности каждой квотной группы в генеральной совокупности.
4. Рассчитать объем выборки из каждой квотной группы.
5. Составить задания анкетерам.

Для определения общего объема квотной выборки не существует строгих математических формул, так, как в случайном отборе. Количество респондентов определяется исходя из потребностей анализа информации. Чем большее число социальных групп будет сравниваться, тем большее число респондентов нужно опросить.

Определив общий объем выборки, мы должны рассчитать число людей, которых нужно отобрать из каждой квотной группы. Для этого следует:

1. Определить долю, которую составляет общий объем выборки в объеме генеральной совокупности (n / N).
2. Умножить эту долю на численность каждой квотной группы.

Процедура расчета объемов выборки из квотных групп абсолютно идентична процедуре расчета объемов выборки из страт.

Приведем пример расчета объема выборки из квотных групп, входящих в генеральную совокупность – совокупность мужчин в возрасте 18–59 лет, проживающих в конкретном районе города. Квотный признак – двухмерный, он сформирован на основе двух признаков. Это возраст и брачное состояние. Число возрастных групп, как видно из таблицы 4, равно **девяти**, а градации брачного состояния укрупнены, их всего **две**. Поэтому количество квотных групп равно восемнадцати: $9 \cdot 2 = 18$. Это мужчины 18–19 лет, состоящие в браке, мужчины 18–19 лет, в браке не состоящие, женщины 20–24 лет, не состоящие в браке, женщины 20–24 лет, состоящие в браке и т. д.

Из каждой квотной группы, входящей в генеральную совокупность, мы должны извлечь выборку, объем которой должен быть заранее рассчитан. Приведем таблицу, отражающую состав генеральной совокупности.

Т а б л и ц а 4

**Распределение мужчин административного района города
по возрасту и брачному состоянию**

Возраст (в годах)	Численность мужчин	Состоит в браке	Не состоит в браке
18–19	6307	141	6166
20–24	12361	2627	9734
25–29	10710	6100	4610
30–34	9392	6718	2674
35–39	8827	6808	2019
40–44	9957	7911	2046
45–49	8882	7167	1715
50–54	8100	6638	1462
55–59	4386	3626	760
Всего	78922	47736	31186

Как видно из данных таблицы 4, объем генеральной совокупности – 78 тысяч 922 человека ($N = 78922$). Если объем выборки запланирован в количестве 500 человек ($n = 500$), то $\frac{n}{N} = \frac{500}{78922} = 0,0063353$. На данное число мы должны умножить численность каждой квотной группы. В результате будет получен объем выборки, соответствующий каждой из 18 квотных групп. Нужно обратить внимание на то, что при делении объема выборки на объем генеральной совокупности нужно *оставлять все знаки после запятой*, избегая округления. Иначе сумма численностей квотных групп будет меньше запланированного числа респондентов, и часть объема выборки будет «потеряна».

Если численность мужчин в возрасте 30–34 года, состоящих в браке, составила в генеральной совокупности 6718 человек, то из данной группы следует отобрать сорок три человека. $1462 \cdot 0,00634 = 42,6$. Мужчин в возрасте 50–54 года, не состоящих в браке, будет отобрано девять человек. ($1462 \cdot 0,00634 = 9,27$.) Результаты расчетов вносятся в таблицу квотной выборки, в которой указывается, сколько представителей каждой квотной группы следует опросить.

В результате осуществления подобных расчетов в каждой квотной группе мы добиваемся *полного соответствия генеральных и выборочных пропорций* по каждому признаку в отдельности и по отдельным сочетаниям двух признаков. Так, в генеральной совокупности доля взятой нами для примера группы мужчин в возрасте 30–34 года, состоящих в браке, составила 8,5 % (6718 из 78922). В выборочной совокупности представителей данной группы будет 43 человека из 500, то есть тоже 8,5 % (табл. 4, 5, 6).

Мы видим, что получившееся число людей, которое нужно отобрать из каждой квотной группы, является дробным. Так как трех десятых человека, или 96 сотых человека, не может быть в принципе, мы должны округлить такого рода числа, а затем проверить, получился ли у нас в сумме изначально заданный объем выборки в размере 500 человек.

Т а б л и ц а 5

**Таблица квотной выборки, полученная в результате расчетов
без округления**

Возраст (в годах)	Численность мужчин	Состоит в браке	Не состоит в браке
18–19	39,96	0,89	39,1
20–24	78,3	16,6	61,7
25–29	67,9	38,6	29,2
30–34	59,5	42,6	16,9
35–39	55,9	43,1	12,8
40–44	63,1	50,1	12,96
45–49	56,3	45,4	10,87
50–54	51,3	42,1	9,3
55–59	27,8	22,9	4,8
Всего	500	302,4	197,6

Т а б л и ц а 6

**Таблица квотной выборки, полученная в результате расчетов
с округлением**

Возраст (в годах)	Численность мужчин	Состоит в браке	Не состоит в браке
18–19	40	1	39
20–24	78	16	62
25–29	68	39	29
30–34	60	43	17
35–39	56	43	13
40–44	63	50	13
45–49	56	45	11
50–54	51	42	9
55–59	28	23	5
Всего	500	302	198

Расчет объемов выборки из квотных групп осуществляется по тому же принципу, что и из страт при пропорциональной стратификации. Однако *квотный и стратифицированный отбор имеют принципиальное различие*: в первом случае это неслучайный отбор, во втором – случайный или вероятностный.

Так, из списка элементов каждой страты осуществляется простой случайный или систематический отбор. В случае же квотной выборки составление списков людей, входящих в квотные группы, не является обязательным, а отбор представителей различных квотных категорий осуществляется без применения таблицы случайных чисел. Полученные в результате расчетов объемы квотной выборки распределяются между анкетерами. Каждому анкетеру следует давать задание на опрос респондентов из самых разнообразных социально-демографических или должностных групп. Это делается с целью увеличения вероятности застать нужного респондента во время посещения квартиры (рабочего места) и снижения числа нерезультативных визитов.

Например, анкетеру № 1 опросить 10 человек:

Женщина 18–29 лет	1
Мужчина 18–29 лет	1
Женщина 30–49 лет	2
Мужчина 30–49 лет	2
Женщина 50–59 лет	1
Мужчина 50–59 лет	1
Женщина 60 лет и старше	1
Мужчина 60 лет и старше	1

Нужно особо отметить, что некоторые выборочные показатели могут приводиться в соответствие с генеральными показателями уже после опроса. При этом объем выборки берется увеличенный, с большим «запасом». Например, анкетерам выдаются списки студенческих групп, число которых явно завышено, а после завершения опроса выборочные пропорции приводятся в соответствие с генеральными пропорциями. Так, квотные задания могут быть выданы по таким признакам, как факультет и курс обучения, а «выравнивание» выборочных и генеральных показателей осуществляется по успеваемости.

Квотная выборка может быть:

1. Одноступенчатой.
2. Представлять процедуру отбора единиц на последней ступени *многоступенчатой* выборки.

Одноступенчатая квотная выборка может быть извлечена из относительно небольшой либо достаточно однородной генеральной совокупности – например, совокупности работников предприятия, студентов города. В данном случае социолог не производит предварительный отбор более крупных единиц, а сразу распределяет по категориям единицы наблюдения, к примеру, работников предприятия, студентов.

Приведем пример использования *одноступенчатой квотной выборки* в процессе мониторинга доступности высшего образования в российских регионах. Объектом исследования были студенты-первокурсники дневных отделений государственных вузов (за исключением обучающихся в творческих вузах и вузах силовых ведомств). Обследованы студенты пяти крупных городов России – Владивостока, Екатеринбурга, Казани, Кемерово и Ульяновска.

В задачи исследования входил сравнительный анализ уровня доступности высшего образования в данных городах и описание общих черт, свойственных данному явлению во всех обследуемых населенных пунктах.

Репрезентативность по квотным признакам обеспечивалась в каждом городе с привлечением данных региональной образовательной статистики. Квотные признаки – вузовская принадлежность и форма обучения (контрактная – бюджетная) – были сопряженными, то есть репрезентативность по форме обучения обеспечивалась в каждом вузе в отдельности. Объем выборки в каждом городе – 500 человек. Всего было обследовано 2500 студентов из 150 академических групп 36 вузов. Нужно особо отметить, что были обследованы все без исключения вузы, входящие в объект исследования⁷.

⁷ См.: Мониторинг региональных проблем доступности высшего образования : аналит. доклад / А. К. Ключев, Е. Л. Могильчак, И. В. Кезина и др. Екатеринбург, 2007. 144 с. (Б-ка журн. «Университетское управление: практика и анализ»).

Отказ от объединения студентов пяти городов в единую выборочную совокупность был принципиальным. Генеральная совокупность, на основе которой была бы сформирована данная выборка, в действительности состоит из пяти различных генеральных совокупностей – студентов Владивостока, Екатеринбурга, Казани, Кемерово и Ульяновска. Только автономный анализ каждого массива анкет позволил не только сравнить данные города, но и выявить то общее, что их объединяет.

Нужно отметить, что основной недостаток квотной выборки – высокая вероятность систематических ошибок, прежде всего вследствие предоставления анкетеру большей свободы, чем в случае вероятностного отбора.

Именно поэтому для уменьшения систематических смещений в отбор вносятся *элементы случайного отбора*, то есть применяется *процедура рандомизации* (от англ. random – случайный. Часто рандомизация рассматривается как процедура случайного отбора. Мы рассматриваем данный термин в узком значении этого слова – как использование случайного отбора на этапах подготовки к формированию квотной выборки). В случае *одноступенчатой производственной* выборки, например при формировании выборочной совокупности студентов в высшем учебном заведении, данная процедура проводится путем случайного отбора студенческих групп из их списков.

Если квотный отбор используется на последней ступени *многоступенчатой выборки*, то рандомизация относится к первым ступеням отбора перед тем как начать собственно квотный отбор. Данная процедура, в случае *территориальной* выборки, осуществляется в следующих формах:

- простой случайный или систематический отбор избирательных участков из их полного списка;
- применение анкетером систематического отбора квартир на данных участках;
- применение маршрутного метода (метода случайного пути) для отбора квартир.

Приведем пример инструкции интервьюеру по отбору квартир при реализации квотной выборки с использованием метода случайного пути.

ИНСТРУКЦИЯ ИНТЕРВЬЮЕРУ

Получив квотное задание и начальный адрес для поиска, Вы должны опросить по этому адресу человека, пол, возраст и образование которого подходит по квоте. В семье можно опрашивать только одного человека.

Если по первому адресу Вы провели интервью, то Вы пропускаете три следующие квартиры (или дома, если дома многоквартирные) и звоните в четвертую. Если по первому адресу Вам не удалось найти человека для интервью, он отказался или никого не было дома, то Вы звоните в следующую квартиру и так до тех пор, пока Вы не найдете нужного респондента. После каждого успешного визита Вы пропускаете три квартиры и звоните в четвертую.

Пройдя полностью один дом, Вы идете в следующий дом по той же стороне улицы. Если на своем маршруте Вы встретите административное здание, пропустите его. Если Ваш путь пересечет улица, поверните направо и продолжайте поиск.

Если Вы после поворотов направо описали полный квадрат и вернулись в исходную точку, перейдите на противоположную сторону улицы и продолжайте маршрут от новой стартовой точки. Если улица, по которой Вы идете, закончилась тупиком, поверните назад и идите, не заходя в дома, до первого перекрестка, затем поверните направо и начните поиск респондентов в первом же жилом доме.

Продолжайте поиск по этому правилу до тех пор, пока Вы не найдете всех респондентов согласно Вашему заданию.

В случае формирования общегородской территориальной выборки более правильной была бы рандомизация путем вероятностного отбора избирательных участков. На каждом из отобранных участков следовало бы осуществить систематический отбор квартир, а в каждой отобранной квартире отобрать респондента в соответствии с квотами.

Квотная выборка является по сути дела единственной разновидностью неслучайного отбора, имеющая определенное математическое обоснование и достаточно высокую эффективность. Очень важным при этом является отбор квотных признаков. Хороший результат достигается в том случае, если используется несколько

Характеристики основных типов одноступенчатых выборочных процедур

Название выборочной процедуры	Требуемая информация	Случайная / неслучайная	Необходимость повторных посещений отсутствующих респондентов	Достоинства	Недостатки
Простая случайная	1. Список всех элементов генеральной совокупности. 2. Таблица случайных чисел	Случайная	Есть	Низкая случайная ошибка	Высокие затраты. Опасность нарушения анонимности
Систематическая (механическая)	Список всех элементов генеральной совокупности	Случайная	Есть	Низкая случайная ошибка	Высокие затраты. Опасность нарушения анонимности
Гнездовая (серийная, кластерная)	1. Полный перечень гнезд, входящих в генеральную совокупность. 2. Таблица случайных чисел	Случайная	Есть	Более низкие затраты на проведение исследования, отсутствие опасности нарушения принципа анонимности	Более высокая случайная ошибка выборок

Название выборочной процедуры	Требуемая информация	Случайная / неслучайная	Необходимость повторных посещений отсутствующих респондентов	Достоинства	Недостатки
Стратифици- рованная (районированная, расслоенная, типическая)	1. Список элементов каждой страты. 2. Таблица случай- ных чисел	Случайная	Есть	Самая низкая слу- чайная ошибка вы- борки	Высокие затраты на проведение иссле- дования, опасность нарушения аноним- ности
Квотная	Численность каж- дой квотной группы в генеральной сово- купности	Неслучайная	Нет	Низкие затраты на проведение иссле- дования	Опасность система- тических ошибок

сопряженных признаков, оказывающих статистически существенное влияние на исследуемые социологом переменные. Чем больше будет таких признаков, тем выше уровень репрезентативности квотной выборки, но больше затраты на проведение опроса.

Приведем обобщающую таблицу (табл. 7), характеризующую рассмотренные выборочные процедуры.

Вопросы и задания для самоконтроля

1. В чем состоит отличие квотной выборки от стратифицированной?
2. Перечислите действия социолога в процессе формирования стратифицированной выборки.
3. Можно ли назвать отбор «первого встречного» на улице случайным?
4. Укажите сходства и различия гнездовой и стратифицированной выборки.
5. Дайте определение ошибки репрезентативности.
6. Укажите основной недостаток квотной выборки.

Раздел 3

МНОГОСТУПЕНЧАТАЯ ВЫБОРКА

По количеству ступеней отбора выборка может быть *одноступенчатой* и *многоступенчатой* (об одноступенчатой и многоступенчатой выборках см. на с. 9).

Нужно отметить, что термин «отбор» означает обязательное *сокращение* числа единиц. Так, на первой ступени из совокупности 200 населенных пунктов могло быть отобрано 50, на второй ступени из нескольких сот избирательных участков, имеющих в каждом отобранном населенном пункте, отобрано тридцать и т. д. На разных ступенях единицы отбора разные: на первой ступени – самые большие по численности, на последней – самые малочисленные. *Тип выборочной процедуры на разных ступенях может быть разным.* Так, например, на первых ступенях может применяться случайный отбор, а на последней – квотный.

Применение многоступенчатой случайной выборки упрощает организацию отбора, но увеличивает ошибку репрезентативности. Каждая ступень вносит свой вклад в ошибку, поэтому *чем больше ступеней, тем ошибка репрезентативности выше.* Проблема отбора *на первой ступени* имеет особое значение, так как ошибки на этой ступени вносят наибольший вклад в суммарную ошибку. Так, опрос большого числа респондентов в небольшом числе населенных пунктов сопровождается значительно большей ошибкой, чем опрос меньшего числа людей в большом числе населенных пунктов.

В рамках одного населенного пункта многоступенчатая вероятностная выборка целесообразна только в случае большой численности населения, причем число ступеней не должно превышать трех.

При формировании общенациональных случайных выборок в первую очередь определяются первичные единицы отбора, то есть те составные части генеральной совокупности, которые будут

отбираться на первой ступени. Это могут быть города, сельские районы, другие административные единицы, входящие в состав территориальных образований. Основные требования к такого рода единицам – возможность составления их полного перечня. Чем меньше размер первичных единиц отбора, входящих в генеральную совокупность и соответственно *больше их численность*, тем меньше будет ошибка выборки на первой ступени, а следовательно, и общая ошибка репрезентативности.

Недопустимым является снижение числа отбираемых единиц до нескольких. При построении общегородской выборки на первой ступени отбора должно отбираться минимум *несколько десятков единиц*. При организации общероссийских опросов нередко отбирается около 100 единиц на первой ступени и около 300 – на второй.

Многоступенчатая случайная выборка со стратификацией единиц первой ступени. Как уже было сказано выше, стратификация – это разделение единиц, входящих в генеральную совокупность, на однородные группы – страты. Основная задача стратификации – добиться, чтобы единицы в каждой страте были похожи друг на друга по исследуемым признакам, к примеру, по ценностным ориентациям. Выяснить, насколько однородными оказались страты по исследуемым признакам, можно только после опроса. Поэтому стратификация как этап, предшествующий извлечению выборки, осуществляется по имеющимся в органах государственной статистики показателям, оказывающим влияние на исследуемые переменные.

Стратификация может быть одномерной, двух или трехмерной и многомерной. В первом случае единицы, входящие в генеральную совокупность, подразделяются на группы по единственному признаку, например, экономико-географической принадлежности. Выделяются единицы, относящиеся к Центральному, Уральскому, Западно-Сибирскому и т. д. регионам.

При *многомерной стратификации* для каждой административной единицы (административный район города или области, малый город областного подчинения и т. д.) в органах государст-

венной статистики берется информация по целому ряду показателей, например, число безработных, доля лиц с высшим образованием, число имеющих доход ниже прожиточного минимума, приходящееся на тысячу населения и т. д. При помощи такого метода многомерной классификации, как кластерный анализ, осуществляемого с помощью компьютерных программ, исследователь получает определенное количество кластеров – групп единиц, отличающихся внутренней однородностью по целой совокупности показателей.

Всероссийский центр изучения общественного мнения (ВЦИОМ) имеет опыт применения пятиступенчатой выборки с *многомерной* стратификацией крупных единиц первой ступени – областей краев, республик.

В основу формирования кластеров были положены следующие характеристики регионов:

1. Доля населения с доходами ниже прожиточного минимума.
2. Отношение среднедушевых доходов к прожиточному минимуму.
3. Отношение среднедушевых расходов к прожиточному минимуму.
4. Отношение инвестиций в основной капитал к валовому региональному продукту.
5. Темпы роста инвестиций в основной капитал. Темпы роста валового регионального продукта (в процентах).
6. Уровень безработицы (на конец года; в процентах от экономически активного населения).
7. Отношение валового регионального продукта на душу населения к прожиточному минимуму.

В результате кластеризации по совокупности *экономических показателей* были выделены 15 кластеров.

В результате кластеризации по *политическим показателям* (на основе итогов последних президентских выборов и выборов в Государственную думу) было сформировано четыре кластера.

В конечном итоге в результате «наложения» данных двух кластеризаций получилось **28 групп регионов**.

К центрам по изучению общественного мнения, использующим *двухмерную стратификацию*, относится «Левада-центр». Приведем основные характеристики из опыта построения выборочной совокупности данным центром⁸.

Тип выборки – четырехступенчатая стратифицированная выборка.

Объем выборки – 1600 респондентов.

Число субъектов РФ, охваченных выборкой, – 45.

Число первичных единиц отбора – 130, в том числе:

Более 1 миллиона человек	13 городов
500 тыс. – 1 миллион человек	14 городов
100 тыс. – 500 тыс. человек	28 городов
До 100 тыс. человек	39 городских населенных пунктов
Сельские районы	36

Всего 94 городских населенных пункта и 36 сельских районов.

1. Первая ступень – отбор населенных пунктов.

На первой ступени выборки в качестве первичных единиц отбора выступали городские населенные пункты и сельские административные районы.

1.1. Предварительная стратификация единиц на первой ступени отбора.

Сначала все первичные единицы отбора разделены на 8 групп в соответствии с их принадлежностью к федеральным округам (Северо-Западный, Центральный, Приволжский, Южный, Северо-Кавказский, Уральский, Сибирский и Дальневосточный).

В каждом из федеральных округов, независимо друг от друга, все первичные единицы отбора разбиты на страты в соответствии с численностью проживающего в них населения:

⁸ Описание процесса формирования многоступенчатой общероссийской выборки с предварительной стратификацией общероссийскими центрами изучения общественного мнения получено на сайтах <http://www.levada.ru>, <http://wciom.ru>.

- 1) города численностью более 1 млн человек;
- 2) города численностью от 500 тыс. до 1 млн человек;
- 3) города численностью от 100 тыс. до 500 тыс. человек;
- 4) городские населенные пункты численностью до 100 тыс. человек;
- 5) сельские населенные пункты.

Таким образом, признак расслоения был *двухмерным*: принадлежность к федеральным округам, сопряженная с принадлежностью к населенным пунктам разного типа.

С учетом особенностей расселения населения в регионах образцовано 36 страт. Для каждой страты подсчитано число постоянно проживающих взрослых жителей и доля этой страты во взрослом населении России.

1.2. Отбор населенных пунктов на первой ступени.

Общий объем выборки (1600 респондентов) распределен между всеми стратами пропорционально численности взрослого населения каждой страты. Число первичных единиц отбора определялось исходя из ограничения на среднее число респондентов в одном городе/сельском районе (7–13 респондентов).

Все города с населением свыше 1 млн человек включены в выборку как *самостоятельные статистические объекты* (генеральные совокупности, из каждой извлекается отдельная общегородская выборка).

В каждой из оставшихся страт случайным методом *с вероятностью, пропорциональной размеру*, были отобраны от 1 до 10 (в зависимости от числа респондентов, пришедшегося на страту) городов/сельских районов. Количество анкет, приходящееся на страту, разделено между отобранными городами/сельскими районами поровну.

Всего в выборку включено 130 городов/сельских районов.

2. Вторая ступень – отбор точек опроса.

В городских населенных пунктах производится простой случайный отбор избирательных участков из числа всех избирательных участков этого населенного пункта. В сельских районах про-

изводится случайный (*с вероятностью, пропорциональной размеру*) отбор сел и поселков городского типа из общего списка всех сел и поселков этого района.

Число отобранных избирательных участков (сельских населенных пунктов/поселков) в городах (сельских районах) определялось исходя из ограничения на среднее число респондентов в одной точке опроса (5–9 респондентов).

В каждой первичной точке отбора отобраны от 1 до 10 (в зависимости от числа респондентов, пришедшегося на город/сельский район) точек опроса. Количество анкет, приходящееся на город/сельский район, разделено между отобранными точками опроса поровну.

Всего в выборку включено 280 точек опроса.

3. Третья ступень – отбор домохозяйств.

При отборе домохозяйств используется случайный маршрутный метод (каждое семнадцатое домохозяйство в районах с многоэтажной застройкой, каждое пятое домохозяйство в районах с индивидуальной застройкой).

4. Четвертая ступень – отбор респондента в домохозяйстве.

Респонденты в домохозяйстве отбираются по методу «ближайшего дня рождения» с контролем половозрастных и полообразовательных квот.

Как мы видим, классический вероятностный отбор респондента из списка членов семьи с использованием таблиц случайных чисел или других подобных таблиц заменяется методом «ближайшего дня рождения». Данный метод выполняет функцию исключения влияния субъективизма анкетера на отбор единиц наблюдения. К тому же нужно учитывать, что в данном случае на последней ступени применяется не случайная выборка, а квотная выборка с предварительной рандомизацией.

Приведем пример расчета объема выборки и ее размещения в процессе построения **областной многоступенчатой выборки** с предварительной стратификацией.

ПРИМЕР

Генеральная совокупность – население области в возрасте 18 лет и старше. Объем выборки – 1000 человек.

Крупные территориальные образования в составе области представляют самостоятельные саморепрезентирующие единицы. При формировании областной (краевой, республиканской) выборки в качестве таких единиц могут выступать областной центр и город областного подчинения с наибольшей численностью населения. Каждая из указанных единиц репрезентирует (представляет) саму себя, то есть выборка из каждого территориального образования формируется таким образом, чтобы результаты опроса можно было распространить на соответствующий город. В данном случае предполагается квотирование не только по социально-демографическим параметрам, но и по району проживания.

Еще раз отметим, что в нашем случае представлены *две* саморепрезентирующие территории.

Два указанных города включают в себя 2 млн человек. В областном центре проживает 1,4 млн человек, во втором по численности городе – 600 тыс.

В целом в двух указанных городах проживает 50 % населения области в возрасте 18 лет и старше. $\frac{2000000}{4000000} = 0,5$. При этом в областном центре проживает 35 % генеральной совокупности, во втором по численности населения городе – 15 %.

Следовательно, из 1000 анкет 35 % приходится на областной центр: $1000 \cdot 0,35 = 350$ анкет. На второй город – $1000 \cdot 0,15 = 150$ анкет.

Таким образом, на две саморепрезентирующихся территории приходится 500 анкет. Следовательно, оставшиеся 500 анкет мы должны распределить на оставшейся территории области.

Общая схема четырехступенчатой выборки *за пределами двух наиболее крупных городов* такова:

1. Отбор **населенных пунктов** из каждой страты.
2. Отбор **избирательных участков** в каждом отобранном населенном пункте (в сельских поселениях сразу отбираются домохозяйства).

3. Отбор **домохозяйств** на каждом отобранном избирательном участке.

4. Отбор **респондента** в каждой отобранной квартире, доме в соответствии с социально-демографическими квотами.

Предварительная стратификация единиц первой ступени одномерная (при проведении реального исследования лучше сделать стратификацию хотя бы двухмерной, например, ввести еще один параметр – принадлежность к административно-управленческим (географическим округам)). Признак расслоения – численность населения населенного пункта. Чтобы рассчитать объемы выборки из каждой страты, используем таблицу 8.

Т а б л и ц а 8

Структура генеральной совокупности

Городские населенные пункты с численностью населения	Численность населения, проживающего в поселениях данного типа	Общее число населенных пунктов в стратах
100–150 тысяч	91990	1
50–99 тысяч	264621	8
30–49 тысяч	609517	13
До 30 тысяч	356377	22
Сельские населенные пункты	677495	340
Всего	2000000	

Так как населенные пункты области на территории *вне* двух самых крупных городов делятся на 5 типов по численности проживающего в них населения, сформировано 5 страт, а из каждой страты необходимо отобрать количество людей, пропорциональное численности элементов страты в генеральной совокупности.

Разделим объем выборки на общую численность населения

в стратах. $\frac{500}{2000000} = 0,00025$.

Чтобы узнать, сколько человек нужно отобрать из каждой страты, мы должны численность каждой страты в генеральной совокупности умножить на 0,00025.

$$0,00025 \cdot 91990 = 22,9975 = 23$$

$$0,00025 \cdot 264621 = 66,15525 = 66$$

$$0,00025 \cdot 609517 = 152,3793 = 152$$

$$0,00025 \cdot 356377 = 89,1 = 89$$

$$0,00025 \cdot 677495 = 169,3738 = 169$$

В связи с тем, что в результате округления объем выборки оказался равен не 500, а 499, прибавим недостающую единицу к самому большому объему выборки из страты – 169. Результаты внесем в четвертый столбец «Объем выборки из страт» таблицы 9.

Т а б л и ц а 9

**Соотношение структуры генеральной
и выборочной совокупности**

№ страты	Городские населенные пункты с численностью населения	Численность населения, проживающего в поселениях данного типа	Процент (в генеральной совокупности)	Объем выборки из страт	Процент (в выборочной совокупности)
1	100–150 тыс.	91990	4,5995	23	4,6
2	50–99 тыс.	264621	13,23105	66	13,2
3	30–49 тыс.	609517	30,47585	152	30,4
4	До 30 тыс.	356377	17,8189	89	17,8
5	Сельские населенные пункты	677495	33,87475	170	34
	Всего	2000000	100	500	100

Из данных таблицы 9 видно, что процент населения, отобранного из страт в выборочную совокупность, совпадает с процентом населения, проживающего в соответствующей страте генеральной совокупности (4,5995 и 4,6 % в первой страте; 13,23105 и 13,2 % во второй страте; 30,47585 и 30,4 % – в третьей страте и т. д.).

Примем число единиц, отбираемых на первой ступени отбора, равным 25. В последнем столбце таблицы 10 укажем число населенных пунктов, которое нужно отобрать из каждой страты, оно должно быть рассчитано пропорционально численности населения в страте. Для этого долю каждой страты в общей численности населения, умножим на 25.

$$0,045995 \cdot 25 = 1,14 \text{ (1 населенный пункт)}$$

$$0,1323105 \cdot 25 = 3,31 \text{ (3 населенных пункта)}$$

$$0,3047585 \cdot 25 = 7,62 \text{ (8 населенных пунктов)}$$

$$0,1783055 \cdot 25 = 4,46 \text{ (5 населенных пунктов)}$$

$$0,3387475 \cdot 25 = 8,47 \text{ (8 населенных пунктов)}$$

Т а б л и ц а 10

Объем выборки из страт

Городские населенные пункты с численностью населения	Объем выборки из страт (человек)	Объем выборки из страт (число обследуемых населенных пунктов)
100–150 тыс.	23	1
50–99 тыс.	66	3
30–49 тыс.	152	8
До 30 тыс.	89	5
Сельские населенные пункты	170	8
Всего	500	25

Так как в первой страте оказался всего один элемент, он и попадает в выборку. В остальных стратах из списка населенных пунктов мы отбираем определенное ранее число населенных пунктов *с вероятностью, пропорциональной численности* людей, проживающих в данных пунктах. На каждый населенный пункт в среднем приходится около 20 человек (см. табл. 11).

Размещение выборки по стратам и населенным пунктам

Численность населения в страте (Генеральная совокупность)	Число отбираемых населенных пунктов из страты	Объем выборки из страты (500 чел.)	Объем выборки в отобранных населенных пунктах
1 страта 91 928 человек	1 городское поселение	Объем выборки из 1-й страты	23 человека
2 страта 264 621 человек	3 1 городское поселение 2 городское поселение 3 городское поселение	Объем выборки из 2-й страты	66 человек 22 человека 22 человека 22 человека
3 страта 609 517 человек	8 1 городское поселение 2 городское поселение 3 городское поселение 4 городское поселение 5 городское поселение 6 городское поселение 7 городское поселение 8 городское поселение	Объем выборки из 3-й страты	152 человека 19 человек 19 человек 19 человек 19 человек 19 человек 19 человек 19 человек 19 человек
4 страта 356 377 человек	4 1 городское поселение 2 городское поселение 3 городское поселение 4 городское поселение	Объем выборки из 4-й страты	89 человек 22 человека 22 человека 22 человека 23 человека

О к о н ч а н и е т а б л . 11

Численность населения в страте (Генеральная совокупность)	Число отбираемых населенных пунктов из страты	Объем выборки из страты (500 чел.)	Объем выборки в отобранных населенных пунктах
5 страта 677 495 человек	8	Объем выборки из 5-й страты	170 человек
	1 село		21 человек
	2 село		21 человек
	3 село		21 человек
	4 село		21 человек
	5 село		21 человек
	6 село		21 человек
	7 село		22 человека
	8 село		22 человека

Таким образом, 1000 анкет были распределены нами между 7 стратами, две из которых – саморепрезентирующие. В остальных 5 стратах опрошенное население не представляет ни саму страту, ни город, в котором проходил опрос. Единственная функция формирования выборки в данных стратах – с д е л а т ь в к л а д в о б щ у ю р е п р е з е н т а т и в н о с т ь п о о б л а с т и в ц е л о м .

Из списка населенных пунктов, входящих в каждую страту, необходимо будет отобрать указанное в таблице число пунктов. Отбор случайный, с в е р о я т н о с т ь ю , п р о п о р ц и о - н а л ь н о й ч и с л е н н о с т и н а с е л е н и я н а с е л е н - н ы х п у н к т о в .

Процедура отбора единиц с вероятностью, пропорциональной их размеру. Отбор единиц с вероятностью, пропорциональной числу входящих в них элементов, используется прежде всего при отборе населенных пунктов на первых ступени многоступен-

чатой территориальной выборки. Это вызвано неравным размером единиц отбора.

Для формирования выборки населенных пунктов:

1) определяется количество обследуемых пунктов (объем выборки на 1-й ступени);

2) определяется шаг отбора путем деления общего числа проживающих во всех населенных пунктах на число обследуемых пунктов (проживающие должны относиться к объекту исследования. Например, если это женщины трудоспособного возраста, то в расчетах используется численность женщин этого возраста);

3) составляется полный список населенных пунктов, входящих в генеральную совокупность (или в каждую из страт, если пункты существенно отличаются по численности населения);

4) рядом с названием каждого населенного пункта указывает-ся количество жителей, входящих в объект исследования;

5) для каждого населенного пункта рассчитываются накопленные (кумулятивные) частоты путем суммирования числа жителей данного пункта с общим числом жителей, проживающих во всех «вышележащих» в списке населенных пунктах;

6) для каждого населенного пункта указывается интервал списочных номеров, соответствующих данному пункту;

7) производится систематический отбор номеров в списках, каждый из отобранных номеров указывает на населенный пункт, где должен быть произведен опрос.

ПРИМЕР

Объектом исследования является население области в возрасте 18 лет и старше.

Общий объем выборки (800 респондентов) распределен между всеми стратами пропорционально численности взрослого населения каждой страты. Одна из страт – это население поселков городского типа, которое составляет в данной области 135 тысяч 200 человек. Общее число поселков – 24. Необходимо отобрать 5 поселков, в которых будет проходить опрос.

Единицей отбора является поселок. Объем выборки на первой ступени был определен равным пяти. Поэтому следует отобрать

из списка 5 чисел, которые укажут, в каких поселках будет проходить опрос. Составим таблицу 12 для формирования выборки на примере отбора из списка поселков городского типа.

Т а б л и ц а 12

**Расчетная таблица для отбора единиц с вероятностью,
пропорциональной их величине**

Поселок	Число жителей	Накопленная частота	Интервал номеров
1	5980	5980	1–5980
2	10386	16366	5981–16366
3	5270	21636	16367–21636
4	7677	29313	21637–29313
5	7955	37268	29314–37268
6	3391	40659	37269–40659
7	3337	43996	40660–43996
8	6767	50763	43997–50763
9	2262	53025	50764–53025
10	3284	56309	53026–56309
11	13133	69442	56310–69442
12	7656	77098	69443–77098
13	6791	83889	77099–83889
14	6790	90679	83890–90679
15	5921	96600	90680–96600
16	4802	101402	96601–101402
17	4518	105920	101403–105920
18	3977	109897	105921–109897
19	3743	113640	109898–113640

Поселок	Число жителей	Накопленная частота	Интервал номеров
20	1841	115481	113641–115481
21	4543	120024	115482–120024
22	4342	124366	120025–124366
23	3510	127876	124367–127876
24	7324	135200	127877–135200

В первом по списку поселке число жителей указанного возраста составляет 5 980 человек. Следовательно, этому поселку соответствуют номера от первого до 5980-го. В поселке 2 живет 10 386 человек. Следовательно, номера жителей этого населенного пункта в нашем списке будут начинаться сразу же после последнего номера поселка 1, то есть с № 5981 и заканчиваться № 16366. Таким образом, рядом с каждым поселком мы указали в таблице тот интервал, к которому относятся номера жителей этого города в нашем списке.

Для того чтобы начать отбор номеров из списка, необходимо определить *шаг отбора*, для этого нужно разделить общее число жителей во всех 24 поселках на то число поселков, в которых будет проходить опрос. Разделив 135 200 на 5, получим 27 040. Это и будет **шаг отбора**.

В случае отбора населенных пунктов из каждой из 5 страт, которые мы описывали ранее в другом примере, шаг отбора во второй страте составил бы 88 207, в третьей – 76 190, в четвертой – 88 150, в пятой (сельские населенные пункты) – 74 166.

Первый номер в нашем списке мы определяем по таблице случайных чисел (см. табл. 1) Первое в первой строке шестизначное число – 661942. Это число и должно указать, какой поселок первым попадет в выборку. Для этого определим, в какой интервал попадает данное число. Из таблицы видно, что оно не попадает ни в один из интервалов, так как общее число людей во всех поселках

меньше 661942, оно составляет всего лишь 135200. Следующие 4 числа также не попадают ни в один из интервалов. Это 892609, 547166, 254551, 567953.

Первое шестизначное число, попавшее в один из интервалов, – 121087. Это интервал 120025–124366. Ему соответствует поселок под номером 22. Это и будет первая единица, которая попадает в выборку. Прибавим к 121087 число 27040 (шаг отбора) и получим 148127. Такого номера опять нет ни в одном интервале. Поэтому начинаем двигаться не вниз, а вверх по таблице. Для этого из нашего номера 121087 вычитаем шаг отбора 27040, получаем 94047. Это число попадает в интервал 90680–96600, соответствующий поселку № 15.

Из 94 047 опять вычитаем 27040, получаем 61007. Это число находится в интервале 56310–69442, соответствующему поселку № 11.

$61007 - 27040 = 33967$ (интервал 37269–40659 – поселок № 6).

$33967 - 27040 = 6927$ (интервал 5981–16366 – поселок № 2)

Таким образом, из 24 поселков мы отобрали необходимые нам 5 поселков – № 22, 11, 15, 6, 2.

Необходимо еще раз отметить, что мы отбирали номера конкретных жителей страты, а, будучи отобранным, этот житель указывал на тот населенный пункт, который и попадал в нашу выборку. (Тот же самый принцип в случае наличия такой возможности было бы целесообразно применять при отборе домохозяйств при формировании многоступенчатой территориальной выборки, когда люди отбираются из избирательных списков с целью определения обследуемых домохозяйств.)

Случайный отбор респондента в семье. Если на последней ступени мы планируем отобрать респондента в семье вероятностным образом, то для этого могут применяться **таблицы Киша** (табл. 13). Для упрощения процедуры отбора в семье иногда в выборку отбирается человек, чей день рождения ближе всего к дате посещения квартиры анкетером.

ПРИМЕР

Т а б л и ц а 13

**Таблицы случайного отбора респондента в семье
(таблицы Киша)**

№ карты	Частота применения	Число членов семьи					
		1	2	3	4	5	6 и более
1	1/6	1	1	1	1	1	1
2	1/12	1	1	1	1	2	2
3	1/12	1	1	1	2	2	2
4	1/6	1	1	2	2	3	3
5	1/6	1	2	2	3	4	4
6	1/12	1	2	3	3	4	5
7	1/12	1	2	3	4	5	5
8	1/6	1	2	3	4	5	6

В приведенной таблице каждая строка является картой, с помощью которой интервьюер может провести отбор респондента. Каждая из восьми карт имеет строго определенную частоту применения, она указана во второй колонке таблицы. Например, если необходимо заполнить 1200 анкет, то по первой карточке респонденты будут отбираться в 1/6 части случаев или при использовании 200 анкет ($1200 : 6 = 200$), по второй карточке – в 1/12 части случаев или при использовании ста анкет ($1200 : 12 = 100$) и т. д.

Придя в квартиру, интервьюер составляет список членов семьи, относящихся к генеральной совокупности, например, лиц избирательного возраста. Затем производится их упорядочивание по заранее определенному исследователем принципу. Таким образом, каждый член семьи получает свой порядковый номер. Это позволяет интервьюеру начать вероятностный отбор человека для опроса по инструкции, используя карточку из таблиц Киша.

ИНСТРУКЦИЯ ИНТЕРВЬЮЕРУ

Для того чтобы определить человека, который будет отвечать на вопросы анкеты, Вам нужно составить список всех людей избирательного возраста. В список нужно включить людей, реально проживающих по данному адресу, не обязательно прописанных здесь. Не включайте в список тех, кто уехал на срок более 6 месяцев или приехал погостить на срок менее 6 месяцев.

Сначала перечислите всех мужчин в возрасте 18 лет и старше от самого старшего до самого младшего с указанием возраста (составление списка может быть осуществлено и по другому принципу – перечисление женщин, а затем мужчин, начиная от самой младшей до самой старшей). Имена называть необязательно, достаточно назвать возраст. Затем женщин в том же порядке – от самой старшей до самой младшей (с указанием возраста). Впишите всех членов семьи в таблицу.

Перечень членов семьи с указанием возраста (составляется интервьюером)

Номер члена семьи	Пол	Возраст (в годах)
1	Мужчина	65
2	Мужчина	42
3	Женщина	38
4	Женщина	20
5	Женщина	18

Карта отбора (дается интервьюеру социологом)

Если число членов семьи равно	Отберите человека под номером
1	1
2	1
3	1
4	1
5	2
6 и более	2

Определите число членов семьи. В левом столбце карты найдите равное ему число. В правом столбце найдите порядковый номер члена семьи, которого нужно опросить.

Если анкетер работает по карте отбора 2 и число членов семьи составило 5 человек, то ему следует опросить респондента № 2, мужчину 42 лет.

Мы рассмотрели различные варианты построения многоступенчатой территориальной выборки со стратификацией единиц первой ступени отбора. Нужно еще раз отметить, что на последней ступени многоступенчатой выборки может быть применен не только «классический» вероятностный отбор респондента при помощи таблиц Киша, но и квотный отбор или отбор с элементами квотирования, как в практике «Левада-центра».

В целом нужно отметить, что существует объективное противоречие между жесткими требованиями, предъявляемыми к формированию выборки статистической теорией, и реальными возможностями исследователя – финансовыми, организационными, а также связанными с влиянием человеческого фактора. Задача социолога – находить компромиссные решения в исследовательской ситуации, минимизировать вероятность появления систематических и случайных ошибок.

Вопросы и задания для самоконтроля

1. Как изменяется ошибка случайной выборки при увеличении числа ступеней отбора?
2. Чем отличается многоступенчатая выборка от одноступенчатой?
3. Как осуществляется предварительная стратификация единиц на 1-й ступени отбора в исследованиях «Левада-центра»?

Раздел 4

СЛУЧАЙНЫЕ И СИСТЕМАТИЧЕСКИЕ ОШИБКИ ВЫБОРКИ

Какие бы типы выборочных процедур ни применялись в эмпирическом социологическом исследовании, конечная задача исследователя – оценить, с какой степенью точности полученные выборочные показатели отражают показатели генеральной совокупности. Поэтому необходимо оценивать размер ошибок репрезентативности.

Об ошибках репрезентативности см. на с. 10–11.

Оценка априорной (доопытной) ошибки может быть выражена лишь в виде вероятностного умозаключения, так как информации о генеральном показателе не может быть получено в принципе. Например, ни в каких статистических справочниках нет информации о мнениях, намерениях всего населения города (страны) в данный момент времени. Поэтому сведения о такого рода явлениях могут быть получены только в результате проведения выборочного, а не сплошного исследования. После проведения опроса по случайной выборке мы можем сделать заключение: с вероятностью в 95 % можно утверждать, что доля придерживающихся определенного мнения колеблется от 26 до 32 % и составляет 29 ± 3 %.

Сделать заключение о доле таких людей с вероятностью в 100 % мы не можем, так как не имеем информации о мнении всех жителей, входящих в генеральную совокупность.

Все ошибки выборки делятся также на *случайные и систематические*. В приведенном нами примере ошибка, которая составила 3 %, является случайной ошибкой.

Случайная ошибка выборки – это ошибка вероятностного (случайного) отбора, которая представляет собой отклонение выборочного показателя от генерального, выраженное в виде разности

выборочной и генеральной доли (средней) и вызванное исключительно несплошным характером обследования. Указанное отклонение *уменьшается с увеличением числа обследуемых единиц.*

Систематическая ошибка – отклонение выборочного показателя от генерального, выраженное в виде разности выборочной и генеральной доли (средней) и вызванное несоответствием способа формирования выборочной совокупности методическим правилам.

1. Случайные ошибки выборки

Случайная ошибка входит в *общую ошибку* выборки как ее составная часть, ее вторая составная часть – систематическая ошибка.

Расчет предельной ошибки *простой случайной выборки*. Предположим, что жители города в возрасте 18 лет и старше составляют один миллион человек. Социолог хочет получить информацию о доле людей в городе, которым нравится смотреть спортивные телеканалы. Однако опросить всех жителей города невозможно, поэтому планируется выборочное исследование с объемом простой случайной выборки, равным 400 единиц. Из таблиц случайных чисел (*Приложение*, табл. 1) было выписано 400 чисел, которые определили номера тех респондентов в списке, которые будут опрошены. В результате опроса выяснилось, что нравится смотреть спортивные каналы 18 % респондентов.

Для отбора номеров респондентов можно было использовать другие строки (столбцы) таблицы случайных чисел. Тогда эти числа были бы другими и, следовательно, в выборку попали бы другие люди. Среди этих людей доля любителей спортивных каналов составила бы не восемнадцать, а, например, 22 %. В третьей выборке эта доля могла оказаться равной 16 % и т. д.

В случае извлечения 100 выборок из той же генеральной совокупности в них были бы зафиксированы самые разнообразные процентные показатели. Лишь часть их в точности совпадала бы с истинной, генеральной долей. Остальные колебались бы вокруг нее с определенным «размахом». При этом в подавляющем большинстве случаев – примерно в девяносто пяти выборках из ста – эта доля отклонилась бы от генеральной не более чем на 5 %.

В реальном исследовании из генеральной совокупности извлекается одна, а не сто выборок. Поэтому величина максимально возможного отклонения выборочного показателя от генерального определяется по специальной формуле. Данное отклонение называется *предельной ошибкой выборки*.

Среднее квадратическое отклонение выборочных долей от генеральной доли называется *средней ошибкой* выборочной доли.

ПРИМЕР

Еще раз напомним, что спортивные каналы нравится смотреть 18 % обследуемых.

Для простой случайной и систематической выборки формула расчета предельной ошибки выборки имеет следующий вид:

$$\Delta = t \cdot \mu, \quad (1)$$

где t – доверительный коэффициент; μ – стандартная (средняя) ошибка выборки. Предельная ошибка выборки – это априорная ошибка. Для ее расчета нужно знать, чему равна стандартная (средняя) ошибка выборки μ . Число выборок, равное 100, приведено нами для примера, из данной генеральной совокупности можно извлечь намного больше выборок, при чем способ отбора мог быть не только простым случайным, но и другим вероятностным. Во втором случае размер отклонения будет другим.

$$\mu = \sqrt{\frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}}, \quad (2)$$

где n – объем выборки; N – объем генеральной совокупности; p – доля лиц в генеральной совокупности, обладающих значением признака, относительно которого рассчитывается ошибка (генеральная доля); $p(1-p)$ — дисперсия качественного альтернативного признака в генеральной совокупности (σ^2).

Нужно принимать во внимание тот факт, что сведения о генеральной дисперсии по исследуемым признакам отсутствуют у социолога. Поэтому при больших объемах выборки генеральная дисперсия $p(1-p)$ (σ^2) заменяется на выборочную дисперсию $w(1-w)$ (s^2).

w – выборочная доля лиц, обладающих значением признака, относительно которого рассчитывается ошибка.

Соотношение выборочной и генеральной дисперсии определяется следующей формулой: $\sigma^2 = s^2 \frac{n}{n-1}$. Как видно, при большом

объеме выборки числитель и знаменатель дроби будут практически равны друг другу, и данная дробь превратится в единицу.

Стандартная (средняя) ошибка простой случайной выборки рассчитывается по формуле (3).

$$\mu = \sqrt{\frac{w(1-w)}{n} \cdot \frac{N-n}{N-1}}, \quad (3)$$

где n – объем выборки; N – объем генеральной совокупности; w – доля лиц в выборочной совокупности, обладающих значением признака, относительно которого рассчитывается ошибка выборочной доли; $w(1-w)$ – дисперсия качественного альтернативного признака в выборочной совокупности.

При большом объеме генеральной совокупности данная формула упрощается и имеет следующий вид:

$$\mu = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}. \quad (4)$$

Величина $w(1-w)$ рассчитывается после опроса. Если в выборочной совокупности нравится смотреть спортивные каналы 18 % опрошенных при объеме выборочной совокупности 400 человек, а генеральной – 1 000 000, то $w(1-w) = 0,18(1-0,18) = 0,18 \cdot 0,82 = 0,1476 = 0,148$; $n = 400$; $N = 1\,000\,000$.

Стандартная ошибка выборки μ составит в данном случае

$$\sqrt{\frac{0,148}{400} \left(1 - \frac{400}{1\,000\,000}\right)} = \sqrt{0,00037 \cdot 0,9996} = 0,019.$$

Коэффициент доверия t определяется исходя из значения доверительной вероятности. Социолог должен решить, какое значе-

ние данной вероятности его устроит. Обычно приемлемым вариантом считается значение вероятности, равное 0,95. В этом случае исследователь будет уверен в своих выводах относительно параметров генеральной совокупности на 95 % из возможных 100 %. Коэффициент в этом случае составит 1,96. На практике при расчетах чаще применяется коэффициент доверия, равный 2, из-за удобства в расчетах. Доверительная вероятность в этом случае увеличивается до 0,954.

Коэффициент доверия, соответствующий доверительной вероятности, определяется по таблице «Значение интеграла вероятностей нормального закона распределения» (см. *Приложение*, табл. 2). При небольших объемах выборки (менее 60) данный коэффициент определяется по таблице «Критические точки распределения Стьюдента» (см. *Приложение*, табл. 3 – для $df = n - 1$, $P = 0,95$.)

В нашем случае значение доверительной вероятности может быть принято равным 0,954, а коэффициент доверия, соответственно, двум. Поэтому $\Delta = t \cdot \mu = 2 \cdot 0,019$. Следовательно, мы можем утверждать с вероятностью 0,954 (954 шанса из тысячи), что доля любителей спортивных каналов в генеральной совокупности составит $0,18 \pm 0,038$ ($18 \pm 3,8$ %). Данный показатель будет находиться в пределах от 14,2 до 21,8 %.

Чтобы достичь требуемой ошибки выборки, необходимо рассчитать ее объем.

Приведем формулу расчета *объема простой случайной выборки*, чтобы выяснить, сколько человек вероятностным образом нам было необходимо отобрать, чтобы достичь 5-процентной предельной ошибки выборки в случае, если выборочная доля равна 18 %. Очевидно, что в реальном исследовании объем выборки рассчитывается до опроса, поэтому выборочная доля не может быть известной до окончания этапа обработки данных

$$n = \frac{Nt^2w(1-w)}{N\Delta^2 + t^2w(1-w)}, \quad (5)$$

где $N = 1\,000\,000$; $\Delta = 0,05$ (± 5 %).

Доверительная вероятность P принята равной 0,954, ей соответствует коэффициент доверия, равный 2.

$$n = \frac{1000000 \cdot 2^2 \cdot 0,18(1 - 0,18)}{1000000 \cdot 0,05^2 + 2^2 \cdot 0,18(1 - 0,18)} = \frac{10000000 \cdot 4 \cdot 0,148}{1000000 \cdot 0,0025 + 4 \cdot 0,148} =$$

$$= \frac{148000 \cdot 4}{2500,592} = \frac{592000}{2500,592} = 236,54 = 237.$$

Из результатов расчетов видно, в целях обеспечения 5-процентной ошибки выборки для доли любителей смотреть спортивные каналы достаточно было отобрать всего лишь 237 человек. Небольшой объем выборки вызван высокой однородностью совокупности: большинство школьников (82 %) относится к одной и той же категории – они *не* являются любителями спортивных каналов. Для процентных показателей, соответствующих *другим переменным*, объем выборки мог оказаться другим.

Объем выборки, достаточный для обеспечения пятипроцентной ошибки относительно *всех признаков*, измеряемых в исследовании, рассчитывается на основе максимально возможного значения выражения генеральной дисперсии качественного альтернативного признака $p(1 - p)$. Оно достигается при $p = 0,5$, когда совокупность распадается на две одинаковые по численности части, одна из которых придерживается определенного мнения или способа поведения, другая – нет. Объем выборки в данном случае будет взят с определенным запасом. В качестве предельной ошибки выборки примем 0,05 ($\pm 5\%$). Расчет должен проводиться по следующей формуле:

$$n = \frac{Nt^2 p(1 - p)}{N\Delta^2 + t^2 p(1 - p)}, \quad (6)$$

где N – объем генеральной совокупности; p – генеральная доля единиц, обладающих значением признака, относительно которого рассчитывается ошибка выборочной доли; t – доверительный коэффициент; Δ – предельная ошибка выборки.

После преобразования получим

$$n = \frac{Nt^2 p(1-p)}{N\Delta^2 + t^2 p(1-p)} = \frac{1}{\frac{\Delta^2}{t^2 \cdot p(1-p)} + \frac{1}{N}}.$$

Мы видим, что выражением $\frac{1}{N}$ при больших N можно пренебречь. Поэтому

$$n = \frac{t^2 p(1-p)}{\Delta^2}. \quad (7)$$

Так как коэффициент доверия t при доверительной вероятности 0,954 составляет 2, а дисперсию качественного альтернативного признака мы принимаем максимально возможной, соответствующей доле 0,5, то формула может быть представлена следующим образом (первый раз в русскоязычной литературе данную формулу привел В. И. Паниотто⁹):

$$n = \frac{t^2 p(1-p)}{\Delta^2} = \frac{2^2 \cdot 0,5(1-0,5)}{\Delta^2} = \frac{4 \cdot 0,25}{\Delta^2} = \frac{1}{\Delta^2}.$$

$$\frac{1}{\Delta^2} = \frac{1}{0,05^2} = \frac{1}{0,0025} = 400.$$

Для простой случайной (а также механической) выборки приводим таблицу (табл. 14) с результатами расчетов объема выборки при заданной предельной ошибке (доверительная вероятность $P = 0,954$).

Используя данную таблицу, можно выбрать устраивающую исследователя предельную ошибку выборки. Затем нужно определить, сколько человек следует случайным образом отобрать и опросить, чтобы обеспечить требуемый размер ошибки по любому из исследуемых признаков. Так, если социолога устраивает 5-процентный

⁹ См.: Паниотто В. И. Качество социологической информации. Киев, 1986.

**Соответствие объема простой случайной выборки
и ее априорных ошибок**

Ошибка выборки (в процентах)	7	5	4	3	2	1
Объем выборки	204	400	625	1111	2500	10000

размер отклонения выборочного показателя от генерального, ему достаточно опросить 400 человек, используя простую случайную или систематическую выборку. Если необходимы более точные результаты, например, 3-процентное отклонение, то следует выбрать объем выборки, равный 1111 человек.

Таким образом, чем больше респондентов мы отберем, используя простую случайную и систематическую выборку, тем меньше будет *случайная* ошибка выборки. Обычно в общественных науках допускается ошибка выборки в размере 5 %.

Нужно особо отметить, что требуемый объем любой разновидности случайной выборки определяется не только задаваемой исследователем случайной ошибкой, но и *потребностями анализа результатов опроса*. Чем на большее число подгрупп разбивается выборка в целях сравнительного анализа, тем больше должен быть объем выборки.

Рассчитаем предельную ошибку еще одной разновидности одноступенчатого случайного отбора – *стратифицированной выборки*.

ПРИМЕР

Генеральная совокупность – совокупность работников медицинских учреждений города, численность которых составляет 10 тысяч человек. Признак расслоения – должность медицинского работника, признак имеет три значения, поэтому образуется три страты:

- врачи;
- медсестры;
- санитарки.

В результате выборочного обследования при помощи стратифицированной выборки выяснилось, что 53 % работников доверяют своему непосредственному руководителю. Объем выборки – 500 человек. Необходимо рассчитать предельную ошибку выборки относительно доли высказавших данное мнение. В результате мы получим интервал, в который попадет *генеральная доля* доверяющих руководству.

Приведем формулу (8) расчета средней ошибки стратифицированной выборки

$$\mu = \sqrt{\frac{s_w^{-2}}{n} \cdot \frac{N-n}{N-1}}. \quad (8)$$

$$s_w^{-2} = \frac{\sum_{i=1}^m w_h (1-w_h) n_h}{n}, \quad (9)$$

где m – число районов (страт); μ – средняя ошибка стратифицированной выборки; s_w^{-2} – средняя из внутрирайонных дисперсий выборочной доли доверяющих руководству (средняя из дисперсий внутри страт); w_h – выборочная доля доверяющих руководству в страте h ; n_h – объем выборки из страты h ; N – объем генеральной совокупности; n – объем выборочной совокупности.

Рассчитаем s_w^{-2} . Для этого построим промежуточную расчетную таблицу (табл. 15).

$$s_w^{-2} = \frac{\sum_{i=1}^m w_h (1-w_h) n_h}{n} = \frac{108,5}{500} = 0,217$$

$$\begin{aligned} \mu &= \sqrt{\frac{s_w^{-2}}{n} \cdot \frac{N-n}{N-1}} = \sqrt{\frac{0,217}{500} \cdot \frac{10000-500}{10000-1}} = \\ &= \sqrt{0,000434 \cdot 0,950} = \sqrt{0,0004123} = 0,02. \end{aligned}$$

$$\Delta = t \cdot \mu = 2 \cdot 0,02 = 0,04.$$

Расчет математического выражения $\sum_i^m w_h(1 - w_h)n_h$
для расчета s_w^{-2}

Страты	Число единиц в страте (генеральная совокупность) N_h	Объем выборки из страт n_h	Число людей, доверяющих руководству в стратах	Выборочная доля доверяющих руководству в стратах w_h	Выборочная дисперсия единиц в стратах $w_h(1 - w_h)$	$w_h(1 - w_h)n_h$
Врачи	5000	250	175	0,70	0,21	52,5
Медсестры	4000	200	80	0,40	0,24	48
Санитарки	1000	50	10	0,20	0,16	8
h	$\Sigma = 10000$ N	$\Sigma = 500$ n	$\Sigma = 265$			$\Sigma = 108,5$

То есть предельная ошибка выборочной доли доверяющих руководству составляет **4 %**.

Рассчитаем процент доверяющих руководству в выборочной совокупности.

$$\frac{265}{500} \cdot 100 \% = 53 \%$$

В генеральной совокупности этот процент составляет $53 \% \pm 4,0$, то есть попадает с вероятностью 0,954 в интервал от $[53 - 4 \%]$ до $[53 + 4 \%]$, то есть от 49 до 57 %.

Если бы мы рассчитали предельную ошибку *простой случайной* выборки на основе формулы (4) стандартной ошибки выборки при тех же пятистах отобранных единицах, то она превысила бы 4 % и составила 4,4 %. Это говорит о том, что стратификация уменьшила ошибку на 0,4 %.

Рассмотрим, как рассчитывается предельная *ошибка гнездовой выборки*.

ПРИМЕР

Простая случайная выборка объемом *30 школьных классов* была извлечена из *всей совокупности 11-х классов города*, число которых *составляет 500*. Школьникам был задан вопрос о том, приходилось ли им посещать тренажерный зал, спортивные секции за последние 12 месяцев.

Определить предельную ошибку выборки и доверительный интервал для доли посетивших тренажерный зал в генеральной совокупности с доверительной вероятностью $P = 0,954$.

Т а б л и ц а 16

Исходные данные для расчета предельной ошибки гнездовой выборки

Классы	Число учеников в классе n_i	Посещают тренажерный зал, спортивные секции		Доля посещающих тренажерный зал, спортивные секции в гнездах w_i
		Да	Нет	
1	25	5	20	0,2
2	25	7	18	0,28
3	30	6	24	0,2
4	30	10	20	0,333333
5	28	7	21	0,25
6	30	6	24	0,2
7	30	5	25	0,166667
8	30	6	24	0,2
9	26	4	22	0,153846
10	25	6	19	0,24
11	30	6	24	0,2
12	28	7	21	0,25

О к о н ч а н и е т а б л . 16

Классы	Число учеников в классе n_i	Посещают тренажерный зал, спортивные секции		Доля посещающих тренажерный зал, спортивные секции в гнездах w_i
		Да	Нет	
13	27	9	18	0,333333
14	24	6	18	0,25
15	30	5	25	0,166667
16	25	5	20	0,2
17	24	6	18	0,25
18	24	8	16	0,333333
19	30	5	25	0,166667
20	30	3	27	0,1
21	26	8	18	0,307692
22	30	5	25	0,166667
23	30	6	24	0,2
24	32	8	24	0,25
25	28	7	21	0,25
26	27	9	18	0,333333
27	30	5	25	0,166667
28	27	9	18	0,333333
29	28	7	21	0,25
30	26	13	13	0,5
Всего	835	199	636	0,238323

Предельная ошибка выборки определяется по формуле $\Delta = t \cdot \mu$, где t – коэффициент доверия, соответствующий доверительной вероятности. В нашем случае она составляет 0,954, ей соответствует коэффициент, равный 2.

Рассчитаем среднюю ошибку гнездовой выборки μ .

$$\mu = \sqrt{\frac{s_w^2}{r} \cdot \frac{R-r}{R-1}}, \quad (10)$$

где R – число гнезд, входящих в генеральную совокупность (объем генеральной совокупности, выраженный в числе гнезд); r – число гнезд, входящих в выборочную совокупность (объем выборки, выраженный в числе гнезд); s_w^2 – межсерийная выборочная дисперсия доли.

В нашем случае это доля посещающих спортивные секции.

Приведем формулу выборочной дисперсии для случая, если гнезда имеют *неравную численность*.

$$s_w^2 = \frac{\sum n_i (w_i - w)^2}{n}, \quad (11)$$

где n – общее число обследованных, входящих в выборочную совокупность (объем выборки, выраженный в числе людей); n_i – число единиц в гнезде i ; w_i – доля посещающих спортивные секции в гнезде i ; w – доля посещающих спортивные секции в целом по выборке.

Покажем, как рассчитывается, выражение в числителе формулы (11).

$$\begin{aligned} \sum n_i (w_i - w)^2 = & 25(0,20 - 0,238)^2 + 25(0,28 - 0,238)^2 + \\ & + 30(0,20 - 0,238)^2 + 30(0,33333 - 0,238)^2 + 28(0,25 - 0,238)^2 + \\ & + 30(0,20 - 0,238)^2 + 30(0,16666 - 0,238)^2 + 30(0,20 - 0,238)^2 + \\ & + 26(0,154 - 0,238)^2 + 25(0,24 - 0,238)^2 + 30(0,20 - 0,238)^2 + \\ & + 28(0,25 - 0,238)^2 + 27(0,33333 - 0,238)^2 + 24(0,25 - 0,238)^2 + \\ & + 30(0,16666 - 0,238)^2 + 25(0,20 - 0,238)^2 + 24(0,25 - 0,238)^2 + \\ & + 24(0,33333 - 0,238)^2 + 30(0,16666 - 0,238)^2 + 30(0,10 - 0,238)^2 + \\ & + 26(0,3077 - 0,238)^2 + 30(0,16666 - 0,238)^2 + 30(0,20 - 0,238)^2 + \\ & + 32(0,25 - 0,238)^2 + 28(0,25 - 0,238)^2 + 27(0,33 - 0,238)^2 + \end{aligned}$$

$$+ 30(0,16666 - 0,238)^2 + 27(0,33333 - 0,238)^2 + 28(0,25 - 0,238)^2 + \\ + 26(0,50 - 0,238)^2 = 5,016742.$$

Результаты расчетов по каждому гнезду и суммарные показатели в столбцах представлены в таблице 17.

Т а б л и ц а 17

Результаты расчетов по гнездам и суммарные показатели

Классы	Число учеников в классе n_i	Доля посещающих тренажерный зал, спортивные секции w_i	$w_i - w$	$(w_i - w)^2$	$n_i(w_i - w)^2$
1	25	0,2	-0,03832	0,001469	0,036716
2	25	0,28	0,041677	0,001737	0,043424
3	30	0,2	-0,03832	0,001469	0,04406
4	30	0,333333	0,09501	0,009027	0,270809
5	28	0,25	0,011677	0,000136	0,003818
6	30	0,2	-0,03832	0,001469	0,04406
7	30	0,166667	-0,07166	0,005135	0,154039
8	30	0,2	-0,03832	0,001469	0,04406
9	26	0,153846	-0,08448	0,007136	0,185545
10	25	0,24	0,001677	0,000003	0,000084
11	30	0,2	-0,03832	0,001469	0,04406
12	28	0,25	0,011677	0,000136	0,003818
13	27	0,333333	0,09501	0,009027	0,243728
14	24	0,25	0,011677	0,000136	0,003272
15	30	0,166667	-0,07166	0,005135	0,154039
16	25	0,2	-0,03832	0,001469	0,036716

Окончание табл. 17

Классы	Число учеников в классе n_i	Доля посещающих тренажерный зал, спортивные секции w_i	$w_i - w$	$(w_i - w)^2$	$n_i(w_i - w)^2$
17	24	0,25	0,011677	0,000136	0,003272
18	24	0,333333	0,09501	0,009027	0,216647
19	30	0,166667	-0,07166	0,005135	0,154039
20	30	0,1	-0,13832	0,019133	0,573998
21	26	0,307692	0,069369	0,004812	0,125115
22	30	0,166667	-0,07166	0,005135	0,154039
23	30	0,2	-0,03832	0,001469	0,04406
24	32	0,25	0,011677	0,000136	0,004363
25	28	0,25	0,011677	0,000136	0,003818
26	27	0,333333	0,09501	0,009027	0,243728
27	30	0,166667	-0,07166	0,005135	0,154039
28	27	0,333333	0,09501	0,009027	0,243728
29	28	0,25	0,011677	0,000136	0,003818
30	26	0,5	0,261677	0,068475	1,780346
Всего	835	0,238323	—	0,1833	5,017242

Проведя подстановку и расчет выражения $n_i(w_i - w)^2$, мы можем определить, чему будет равна дисперсия.

$$s_w^2 = \frac{\sum n_i(w_i - w)^2}{n} = \frac{5,017}{835} = 0,006.$$

В нашем случае

$$R = 500$$

$$r = 30$$

$$\begin{aligned}\mu &= \sqrt{\frac{s^2}{r} \cdot \frac{R-r}{R-1}} = \sqrt{\frac{0,006}{30} \cdot \frac{500-30}{500-1}} = \\ &= \sqrt{0,0002 \cdot 0,94} = \sqrt{0,000188} = 0,0137.\end{aligned}$$

Рассчитаем предельную ошибку выборочной доли посещающих спортивные сооружения:

$$\Delta = t \cdot \mu = 2 \cdot 0,0137 = 0,0274.$$

В процентах эта ошибка составит 2,74 %.

В целом по выборочной совокупности посещают тренажерный зал, спортивные секции 199 человек. Всего опрошено 835 человек (см. табл. 17).

Значит, выборочная доля школьников, посещающих тренажерный зал, спортивные секции, в выборке составила $199 : 835 = 0,238$, в процентах – 23,8.

Следовательно, генеральная доля составит $0,238 \pm 0,0274$, то есть попадет в интервал от 0,2106 до 0,2654 или от 21,06 до 26,54 %.

Мы видим, что, несмотря на маленький объем выборки (30 гнезд), ее ошибка оказалась небольшой. Это вызвано тем, что классы очень похожи друг на друга по степени «охвата» организованными спортивными занятиями.

Если мы проведем расчет предельной ошибки *простой случайной выборки* с объемом 835 человек, то она составит не 2,74, а 2,68 %. То есть *простая случайная выборка по сравнению с гнездовой дает в данном случае выигрыш в репрезентативности*, хотя и не очень большой.

В целом априорная ошибка гнездовой выборки выше, чем простой случайной и тем более стратифицированной. Это объясняется тем, что при одинаковом значении дисперсии (числитель

в формуле ошибки), объем выборки (знаменатель) всегда меньше. Дело в том, что в случае простого случайного и стратифицированного отбора объем выборки – это число респондентов, в случае гнездовой выборки – число гнезд. Если сложить численности людей, входящих в гнезда, то это число, естественно, окажется больше количества самих гнезд.

В случае, если гнезда, например школьные классы, существенно различаются по какому-либо свойству, например по принадлежности к обычным школам или лицеям (гимназиям), производится их стратификация.

Стандартная (средняя) ошибка гнездовой выборки с предварительной стратификацией рассчитывается по следующей формуле:

$$\mu = \sqrt{\frac{\overline{s_k^2}}{r} \cdot \frac{R-r}{R-1}}, \quad (12)$$

где $\overline{s_k^2}$ – межсерийная дисперсия доли в k -й страте; R – число серий в генеральной совокупности; r – число серий в выборочной совокупности.

Межсерийная дисперсия в данном случае рассчитывается по следующей формуле:

$$\overline{s_k^2} = \frac{\sum_{k=1}^m r_k s_k^2}{\sum_k r_k}, \quad (13)$$

где r_k – число серий, отобранных в k -й страте (районе); s_k^2 – межсерийная дисперсия доли в k -й страте (районе); m – число страт (районов).

Предельная ошибка выборки рассчитывается по уже известной по другим типам отбора, формуле $\Delta = t \cdot \mu$.

Расчет требуемого *объема гнездовой выборки* можно было бы проводить по стандартной формуле, полученной на основе формулы предельной ошибки гнездовой выборки, где σ^2 обозначает межсе-

рийную дисперсию в *генеральной совокупности*. Однако мы не можем знать, чему равно значение данной дисперсии.

$$r = \frac{t^2 \sigma^2 R}{\Delta^2 (R - 1) + t^2 \sigma^2}. \quad (14)$$

Если принять *дисперсию максимально возможной* и получить объем выборки с некоторым запасом, как мы делали применительно к простой случайной выборке, то данный объем будет очень большим и поэтому нереализуемым на практике. Так, например, при объеме генеральной совокупности, равным 100 единицам, требуемое число обследуемых гнезд будет равно восьмидесяти или составит 80 % от ее объема.

Ясно, что в реальном исследовании генеральная дисперсия вряд ли окажется максимальной, но каким будет ее конкретное значение, мы не можем определить. Единственное, что можно сделать – принять решение о требуемом объеме выборки и оценить его целесообразность. Это делается после обработки данных и получения информации о выборочной дисперсии. Если предельная ошибка выборки, в формулу которой мы подставили выборочную дисперсию, составит не более 5 %, то это означает, что наше решение было правильным.

Для приблизительной оценки требуемого объема выборки следует обратить внимание на расчет выражения, входящего в формулу (10) ошибки гнездовой выборки

$$\frac{R - r}{r(R - 1)}. \quad (15)$$

Чем ближе полученная в результате расчетов доля к нулевому значению, тем меньше будет ошибка выборки. В целом число гнезд должно быть не менее 30.

Мы привели пример расчета ошибок случайной выборки, которые являются *одноступенчатыми*.

Приведем пример расчета ошибки *двухступенчатой случайной выборки*.

Применяемая статистиками формула предполагает, что «крупные» единицы отбора на первых ступенях выборки имеют равную численность. Нужно отметить, что естественные или административные единицы редко имеют абсолютно равное число элементов. Например, это касается *избирательных участков*. Однако в связи с тем, что различия в численности данных участков нельзя назвать существенными, на рассчитываемой ошибке выборки это практически не сказывается.

На первых ступенях случайной выборки, нередко с целью выравнивания численности единиц отбора, некоторые из них объединяют. В других ситуациях при предварительной стратификации данных единиц формируют страты с равной численностью населения. Затем в каждой из равных по численности страт производится случайным образом отбор равного числа административных единиц. Таким способом социологи добиваются «выравнивания» числа элементов в единицах отбора.

Нужно отметить, что многоступенчатый отбор целесообразен прежде всего при использовании территориальной, а не производственной выборки, данный пример приведен чисто в учебных целях, его цель — освоить технологию расчета, поэтому объемы выборки взяты небольшие (приведен адаптированный к социологическим задачам вариант расчетов в книге У. Кокрен¹⁰).

ПРИМЕР

Из 160 производственных бригад, численностью 9 человек в каждой, случайным образом отобрали 40. Затем в каждой из 40 бригад по таблице случайных чисел отобрали для обследования 3 человека из 9. Выяснилось, что в **22 бригадах** не зафиксировано ни одного рабочего, имеющего дисциплинарные взыскания, в **11 бригадах** дисциплинарные взыскания имел один рабочий из трех, в **четыре**х — два, в **трех бригадах** — трое рабочих из трех. То есть в бригадах первой группы взыскания имеет треть рабочих, во второй — две трети, в третьей группе взыскания имеют все рабочие.

¹⁰ См.: Кокрен У. Методы выборочного исследования / под ред. А. Г. Волкова. М., 1976. С. 299–301.

Нужно оценить долю рабочих, имеющих дисциплинарные взыскания и ее предельную ошибку с доверительной вероятностью $P = 0,95$.

Таким образом, имеем двухступенчатую случайную выборку. На первой ступени отобрано 40 бригад из 160, на второй ступени – в каждой из 40 бригад отобрано 3 человека из 9.

$N = 160$; $n = 40$ (первая ступень); $M = 9$; $m = 3$ (вторая ступень).

Всего отобрано 120 рабочих (в каждой из 40 бригад по 3 человека).

По всей выборочной совокупности дисциплинарные взыскания имеет 28 рабочих. В бригадах первой группы – 0, во второй группе – 11 человек (в каждой из одиннадцати бригад по одному человеку), в третьей группе – 8 (в каждой из четырех бригад по два человека), в четвертой группе – 9 (в каждой из трех бригад по три человека).

$0 + 11 + 8 + 9 = 28$. То есть 28 рабочих, имеющих дисциплинарные взыскания, составляют от 120 отобранных 0,233, или 23,3 %.

$$w = \frac{28}{120} = 0,233.$$

Формула стандартной ошибки двухступенчатой выборки приведена ниже:

$$\mu = \sqrt{\frac{1-f_1}{n(n-1)} \sum_{i=1}^n (w_i - w)^2 + \frac{f_1(1-f_2)}{n^2(m-1)} \cdot \sum_{i=1}^n w_i(1-w_i)}, \quad (16)$$

где w_i – доля имеющих взыскания в бригаде; w – доля имеющих взыскания во всей выборке.

Введем следующие обозначения:

$$f_1 = \frac{n}{N};$$

$$f_2 = \frac{m}{M};$$

$$f_1 = \frac{n}{N} = \frac{40}{160} = \frac{1}{4};$$

$$f_2 = \frac{m}{M} = \frac{3}{9} = \frac{1}{3};$$

$$1 - f_1 = 1 - \frac{n}{N} = 1 - \frac{1}{4} = \frac{3}{4};$$

$$1 - f_2 = 1 - \frac{m}{M} = 1 - \frac{3}{9} = 1 - \frac{1}{3} = \frac{2}{3}.$$

Рассчитаем $\sum_i^n w_i(1 - w_i)$ и $\sum_i^n (w_i - w)^2$.

В связи с тем, что в ряде бригад число имеющих дисциплинарные взыскания является одинаковым, вышеприведенные выражения $w_i(1 - w_i)$ и $(w_i - w)^2$ мы будем суммировать не по каждой бригаде, а по группам бригад (табл. 18).

Т а б л и ц а 18

**Исходные данные для расчета стандартной ошибки
двухступенчатой случайной выборки для доли рабочих,
имеющих дисциплинарные взыскания**

Число имеющих взыскания в бригаде	Число бригад	Доля имеющих взыскания в бригаде w_i
0	22	0
1	11	$1/3 = 0,333333$
2	4	$2/3 = 0,666667$
3	3	$3/3 = 1,0$
	40	

$$\begin{aligned}\sum_i^n w_i(1 - w_i) &= 22(0 \cdot (1 - 0)) + 11\left(\frac{1}{3} \cdot \left(1 - \frac{1}{3}\right)\right) + 4\left(\frac{2}{3} \cdot \left(1 - \frac{2}{3}\right)\right) + \\ &+ 3(1 - 1) = 22 \cdot 0 + 11 \cdot \frac{2}{9} + 4 \cdot \frac{2}{9} + 3 \cdot 0 = 0 + 11 \cdot 0,222 + 4 \cdot 0,222 + 0 = \\ &= 2,442 + 0,888 = 3,333.\end{aligned}$$

$$\begin{aligned}\sum_i^n (w_i - w)^2 &= 22(0 - 0,233)^2 + 11(0,333 - 0,233)^2 + \\ &+ 4(0,666 - 0,233)^2 + 3(1 - 0,233)^2 = \\ &= 22 \cdot 0,054289 + 11 \cdot 0,010067 + 4 \cdot 0,188067 + 3 \cdot 0,588289 = \\ &= 1,194358 + 0,1107346 + 0,7522671 + 1,764867 = 3,822.\end{aligned}$$

Теперь рассчитаем *стандартную* (среднюю) ошибку доли рабочих, имеющих дисциплинарные взыскания:

$$\begin{aligned}\mu &= \sqrt{\frac{1 - f_1}{n(n - 1)} \sum_{i=1}^n (w_i - w)^2 + \frac{f_1(1 - f_2)}{n^2(m - 1)} \cdot \sum_{i=1}^n w_i(1 - w_i)} = \\ &= \sqrt{\frac{3 \cdot 3,822}{4 \cdot 40 \cdot (40 - 1)} + \frac{2 \cdot 3,333}{4 \cdot 3 \cdot 40^2 \cdot (3 - 1)}} = \sqrt{0,00201} = 0,045.\end{aligned}$$

Рассчитаем *предельную* ошибку доли рабочих, имеющих дисциплинарные взыскания:

$$\Delta = 1,96 \cdot 0,045 = 0,088.$$

Следовательно, в генеральной совокупности (на данном заводе) доля рабочих, имеющих дисциплинарные взыскания, составляет $23,3 \pm 8,8 \%$. Ошибка оказалась очень большой. Число единиц на первой ступени отбора оказалось недостаточным, поэтому его нужно было увеличить, отобрав более чем 40 бригад.

Подводя итог анализу способов расчета предельной ошибки случайной выборки различных типов, можно сказать, что данная ошибка имеет две главные составляющие:

- выборочная дисперсия доли (средней);
- объем выборки (в том числе количество крупных единиц в случае многоступенчатого отбора).

Чем более похожими друг на друга по исследуемому признаку будут единицы отбора и чем большее число этих единиц мы отберем, тем ошибка выборки будет меньше.

Нужно отдельно отметить, что характер используемых выборочных процедур должен приниматься во внимание при *компьютерной обработке данных*. Так, при гнездовой выборке гнездовая принадлежность должна представлять дополнительную переменную. Поэтому на каждой анкете должен быть указан номер гнезда, к которому принадлежит респондент, который ее заполнил. То же самое касается единиц отбора на первых ступенях выборки, например, избирательных участков.

2. Систематические ошибки выборки

Систематические ошибки выборки отличаются от случайных своими источниками. Случайные ошибки возникают вследствие несплошного характера обследования и уменьшаются с увеличением объема случайной выборки. Систематические ошибки характерны как для случайных, так и неслучайных видов отбора и не уменьшаются при увеличении объема выборки.

Источником систематических ошибок является несоответствие способа формирования выборочной совокупности методическим правилам.

Это несоответствие может быть вызвано действиями социолога, интервьюера или респондента.

Источники систематических ошибок случайной выборки

Требования, предъявляемые к вероятностной выборке, заключаются прежде всего в обеспечении всем элементам генеральной совокупности *равных шансов попадания в выборку*. Этому могут препятствовать на этапе реализации проекта выборки несколько причин:

1. Использование неполной основы выборки (действия социолога).
2. Отступление от правил случайного отбора респондента (действия анкетера).
3. Невозможность вступить в контакт с респондентом (источник ошибки – респондент).

Первая причина возникновения систематических смещений в процессе осуществления случайной выборки может быть проиллюстрирована примером неправильного прогноза результатов голосования американских избирателей в 1936 году. Данные опроса были опубликованы в журнале *Literary Digest* и предсказывали победу А. Ландона над Ф. Рузвельтом с преимуществом в 15 %. В действительности же он проиграл.

Причиной неправильного прогноза послужило использование неполной основы выборки, в которую были включены не все потенциальные избиратели, а только те, кто значился в телефонных справочниках и регистрационных списках автовладельцев. Владельцы телефонов и автомобилей чаще были представителями обеспеченных слоев населения, симпатизировавших Ландону.

Институт изучения общественного мнения Дж. Гэллага дал правильный прогноз результатов голосования, им была применена квотная выборка, представлявшая все типы избирателей в нужных пропорциях – как богатых, так и бедных.

Вторая причина связана с действиями интервьюера как источника систематических ошибок.

Интервьюер вместо случайным образом отобранного респондента может опросить другого члена семьи, который на момент опроса оказался дома и дал согласие ответить на вопросы анкеты.

Возможна также замена анкетером в процессе рандомизации случайным образом отобранной квартиры, где отсутствовали на момент посещения хозяева, другой квартирой.

Систематические ошибки, вносимые интервьюером, могут быть результатом *неправильного принятия решения социологом* об использовании вероятностной выборки. Так, при ограниченном ресурсе времени вероятностная выборка, требующая повторных посе-

щений респондентов, провоцирует интервьюера на замену труднодоступных респондентов легкодоступными.

Третья причина связана с наличием среди населения недоступных и труднодоступных единиц.

Это люди 1) часто находящиеся за пределами постоянного места жительства, их невозможно застать дома даже после 3–4 посещения; 2) неспособные отвечать на вопросы анкеты вследствие болезни или преклонного возраста; 3) категорически отказывающиеся от участия в опросе.

В связи с наличием последней категории людей при любых видах отбора – как случайных, так и неслучайных – *необходимо фиксировать число отказов от участия в опросе*. Высокий процент отказавшихся не позволяет рассматривать полученную в результате опроса информацию как надежную.

В данном случае всю генеральную совокупность можно рассматривать как состоящую из двух слоев. Результаты исследования представляют собой информацию только по первому слою, то есть по людям, которые реализовали возможность принять участие в опросе. Элементы второго слоя генеральной совокупности оказываются не представленными в выборке вообще.

Величина систематического смещения, *вызванного наличием недоступных и труднодоступных единиц*, нарастает 1) с увеличением их доли; 2) с увеличением различий между упомянутыми выше слоями по исследуемым признакам. Если данные слои не отличаются друг от друга, например по намерению приобрести автомобиль в ближайшие полгода, то систематическая ошибка по показателю «доля намеревающихся приобрести автомобиль» будет равна нулю.

Значение систематической ошибки выборки могло бы быть рассчитано только при наличии информации о распределении исследуемых признаков как среди опрошенных, так и среди тех, кого не удалось опросить. Тогда ошибка была бы равна произведению двух величин: 1) доли не принявших участие в опросе; 2) значения *разности долей* людей, обладающих данным значением признака в двух рассматриваемых нами слоях (процедура расчета система-

тической ошибки и описание методов компенсации смещений приведены в соответствии со статьей Г. Н. Сотниковой¹¹).

Мы можем рассчитать *максимально возможное* значение систематической ошибки. Для этого нужно принять в качестве указанной выше *разности долей* максимально возможное значение. Это значение будет *равно единице* или выраженное в процентах – 100 % (в одном слое придерживается определенного мнения 0 %, в другом – 100 %). Поэтому систематическая ошибка составит величину, равную произведению единицы и доли неопрошенных. Следовательно, максимально возможная систематическая ошибка равна доле не принявших участие в опросе.

Для уточнения значения ошибки выборки с учетом присутствия в генеральной совокупности труднодоступных единиц применяются следующие методы:

1. Проводится опрос людей, с которыми удалось установить контакт после первого визита. Затем из совокупности необследованных извлекается малая случайная выборка, и в ней изучаются все элементы без исключения.

2. Выборочная совокупность делится на группы людей, с которыми удалось установить контакт с первого, второго, третьего и т. д. раза. Затем сравниваются их характеристики, измеряемые в анкете. Это позволяет оценить возможность систематических смещений при разовом посещении респондентов и сравнить параметры людей с разной вероятностью контакта.

Оба вышеприведенных метода способствуют получению информации о возможных систематических смещениях, вызванных наличием «подвижных» респондентов. Группа же людей, отказывающихся от участия в опросе, в лучшем случае может быть изучена лишь на уровне социально-демографических параметров.

Нужно отметить, что отступление от принципов случайной выборки неизбежно в эмпирическом социологическом исследовании.

¹¹ См.: Сотникова Г. Н. Методы контроля и компенсации смещений от недо-ступных единиц в выборочном социологическом исследовании // Применение математических методов и ЭВМ в социологических исследованиях. М., 1982. С. 102–117.

Следовательно, неизбежны и систематические ошибки и задача исследователя состоит в их минимизации и оценке.

Систематические ошибки квотной выборки

Квотная выборка автоматически обеспечивает соответствие генеральных и выборочных показателей *только по квотным признакам*. Систематические смещения по *исследуемым характеристикам* могут возникать вследствие того, что отбор из самих квотных групп осуществляется интервьюером с определенной степенью свободы. Процедура данной разновидности выборки предполагает *однократное* посещение респондента. Четкая фиксация отказов от участия в опросе также затруднена. Интервьюер более свободен в своих действиях по сравнению со случайным отбором. Поэтому более вероятно его обращение к людям, которые проявляют интерес к изучаемой в исследовании проблематике, что увеличивает систематическую ошибку выборки.

Источники систематических ошибок квотного отбора

1) использование социологом устаревших данных при расчете объемов выборки из квотных групп;

2) замена анкетером респондента с нужным сочетанием социально-демографических характеристик на другого члена семьи, трудового коллектива, с которым легче установить контакт;

3) нарушение правил рандомизации квотного отбора. Так, к примеру, в процессе применения метода случайного пути возможна замена анкетером труднодоступных участков пути на легкодоступные.

Для снижения вероятности отступления интервьюера от правил квотного отбора важно создать ему условия для быстрой и эффективной работы.

В процессе реализации территориальной выборки с целью *повышения эффективности посещений* интервьюеру может быть передана информация о том, какие категории респондентов проще застать дома в то или иное время; о коэффициентах отсутствия различных социально-демографических групп респондентов.

В квартире в этом случае опрашивается в первую очередь тот член семьи, который относится к социально-демографической

группе с самым высоким коэффициентом отсутствия. Заполняя сначала самые труднодоступные квотные группы, а в последнюю очередь легкодоступные, интервьюер получает возможность с меньшими затратами осуществить опрос в соответствии с планом выборки.

Естественно, чтобы получить сведения, передаваемые анкетеру, необходимо проведение методических исследований.

Апостериорный контроль ошибок репрезентативности

Наиболее просто определить, есть ли систематические смещения в *квотной выборке*, можно, вводя в анкету контрольные признаки, данные о распределении которых в генеральной совокупности есть у исследователя, но которые не использовались при проектировании выборки. Так, например, если в качестве квотных признаков использовались пол, возраст и образование, то контролировать репрезентативность можно по такому признаку, как размер семьи, тип занимаемого жилого помещения, брачное состояние, род занятий и другим имеющимся в органах статистики показателям. Может быть использован контроль размера ошибки по такому показателю, как процент участвовавших в последних президентских выборах или выборах в Государственную думу, местные органы власти. Особое значение при изучении некоторых тем имеет контроль по принадлежности респондентов к группам домохозяйств с разным числом членов.

Для определения наличия систематических смещений в результате *случайного отбора* иногда достаточно проверить соответствие выборки и генеральной совокупности по социально-демографическим параметрам. Однако при стратифицированной выборке это должны быть показатели, отличные от признаков расслоения.

Контроль по такому социально-демографическому признаку, как возраст, должен осуществляться с использованием шкал с достаточно большим числом градаций, которые к тому же должны соответствовать данным органов статистики. Если, например, органы статистики используют такие возрастные градации, как 20–29 лет, 30–39, 40–49 и т. д., а в анкете применяются градации 22–31, 32–41 и т. д., то сравнить показатели генеральной и выборочной совокуп-

ности будет невозможно. Не следует чрезмерно укрупнять последнюю градацию. Если последний вариант ответа на вопрос о возрасте «60 лет и старше», то невозможно будет определить ошибку по принадлежности к таким возрастным категориям, как 70-летние, 80-летние или к укрупненной категории «70 лет и старше». Если люди преклонного возраста будут в меньшей степени представленными в выборке, то за счет большего представительства младших возрастных групп общая доля людей 60 лет и старше может оказаться совпадающей с генеральным показателем. Это скроет реальное систематическое смещение.

Нужно отметить, что введенные в анкету дополнительные вопросы могут рассматриваться всего лишь как *косвенные индикаторы* наличия выборочных ошибок. Чем больше таких индикаторов, тем выше вероятность выявления смещений и по исследуемым признакам – приверженности населения различным взглядам, поведению, установкам и т. д. В качестве общей ошибки выборки принимается максимальная из ошибок всех проверяемых признаков.

Нужно помнить, что ***основная опасность при формировании выборочной совокупности – не ошибки выборки, а отсутствие информации у исследователя об их наличии и размере.***

Ремонт выборки

Ремонт выборки – уравнивание выборочных и генеральных характеристик. Необходимость ремонта вызвана наличием как систематических, так и случайных смещений в выборке.

Для выявления смещений производится сравнение выборочных показателей с известными социологу генеральными показателями. Как правило, это социально-демографические параметры. Единственный способ избежать ремонта выборки – провести дополнительный опрос недостающих категорий респондентов. Например, если в выборке недостает мужчин в возрасте 20–29 лет, следовало бы дать анкетерам задание на дополнительный опрос респондентов с данными характеристиками. В случае невозможности в нужные сроки провести такой опрос принимается решение осуществить ремонт выборки.

Прежде чем приступить к данной процедуре, необходимо выявить влияние признаков, по которым производится коррекция, на исследуемые характеристики. Если это влияние слабое, то ремонт выборки не даст эффекта. Нужно отметить, что нет гарантии, что выравнивание по социально-демографическим параметрам позволит ликвидировать систематическую ошибку, вызванную наличием в генеральной совокупности труднодоступных респондентов.

Перевзвешивание применяется после введения информации, полученной из анкет, в компьютер. Для того чтобы осуществить процедуру перевзвешивания, необходимо каждому значению признака присвоить свой коэффициент, который называют весовым коэффициентом или коэффициентом перевзвешивания.

Он рассчитывается как *частное от деления генеральной доли (процента) людей, которые обладают данным значением признака, на выборочную долю (процент)*.

Если количество респондентов, обладающих определенным значением признака, *избыточно*, то весовой коэффициент будет равен числу, которое меньше единицы. В случае *дефицита* анкет с данным значением признака им присваивается коэффициент больше единицы.

Процедура перевзвешивания осуществляется при помощи специальных программ обработки статистической информации (например, SPSS, Statistica, Vortex и др.) Для этого в компьютер должны быть введены соответствующие коэффициенты.

Приведем пример перевзвешивания по признаку «пол» и проследим влияние этого перевзвешивания на исследуемый признак «намерение голосовать за кандидата X». Объем выборки составил 1000 человек. В таблице 19 представлены выборочные и генеральные показатели для мужчин и женщин, выраженные в процентах, а также их численность в выборке. Из данных таблицы видно, что коэффициент перевзвешивания (весовой коэффициент) для мужчин (1,5) получен путем деления процента в генеральной совокупности на процент в выборочной совокупности. То есть 45 разделили на 30. Для женщин произвели ту же самую процедуру: 55 разделили на 70 и получили 0,786.

Нужно отметить, что исследователю достаточно рассчитать весовой коэффициент и осуществить определенные манипуляции на компьютере. В результате компьютер представит готовое распределение всех исследуемых признаков в отремонтированном массиве.

Т а б л и ц а 19

**Данные для расчета весового коэффициента
по признаку «пол»**

Пол	Процент в выборке	Процент в генеральной совокупности	Коэффициент перевзвешивания	Численность в выборке	Численность в выборке после ремонта
Мужчины	30	45	1,500	300	450
Женщины	70	55	0,786	700	550
Всего	100 %	100 %	—	1000	1000

Основную часть расчетов, приведенных ниже, мы используем с целью прояснения содержания процедуры ремонта, имитируя ремонт выборки без помощи компьютера. Для реализации указанной процедуры следует осуществить:

1. Расчет весовых коэффициентов. Весовые коэффициенты определялись путем деления процентных показателей в генеральной совокупности на выборочные показатели. Для мужчин 45 делилось на 30 (1,5), для женщин – 55 на 70 (0,786).

2. Умножение каждого из коэффициентов на численность соответствующей группы в выборке. В результате получается численность групп в отремонтированном массиве. Так как каждому мужчине присваивался коэффициент равный 1,5, то численность мужчин в выборке после ремонта составила 450 человек ($300 \cdot 1,5 = 450$) Женщинам присваивался коэффициент 0,786. Поэтому их количество равно пятистам пятидесяти ($700 \cdot 0,786 = 550$).

Посмотрим, как изменилось процентное распределение по признаку «Намерение голосовать за кандидата X» после ремонта выборки (табл. 20). Мужчины и женщины отличаются друг от друга

по данному признаку. Так, среди мужчин данное намерение высказало 50 %, а среди женщин – 70 %. Поэтому изменение соотношения мужчин и женщин в выборке неизбежно должно сказаться на распределении анализируемого признака.

Т а б л и ц а 20

**Показатели распределения признака
«Намерение голосовать за кандидата Х» до ремонта выборки**

Намерение голосовать за кандидата Х	% (мужчины)	Абсолютные показатели (чел.)	% (женщины)	Абсолютные показатели (чел.)	Всего (чел.)
Есть	50	150	70	490	640 (64 %)
Нет	35	105	20	140	245 (24,5 %)
Не знаю	15	45	10	70	115 (11,5 %)
Всего	100 %	300	100 %	700	1000

Как видно, в неотремонтированном массиве будет голосовать за кандидата Х 640 человек из 1000, или 64 %, не будет – 245 человек (24,5 %), не приняли решения 115 человек (11,5 %).

Для получения распределения исследуемого признака в отремонтированном массиве без использования компьютера следовало бы произвести следующие действия:

1. Умножение доли желающих голосовать за кандидата Х среди мужчин на численность мужчин в отремонтированном массиве (0,5 умножаем на 450, получаем 225 желающих голосовать за данного кандидата).

2. Умножение доли желающих голосовать за кандидата Х среди женщин на численность женщин в отремонтированном массиве (0,7 умножаем на 550, получаем 385 желающих голосовать за данного кандидата).

3. Складываем 225 и 385. Получаем 610 желающих голосовать за данного кандидата в отремонтированном массиве.

4. Определяем долю желающих голосовать за кандидата X в отремонтированном массиве. 610 делим на 1000. Получаем 0,61, или 61%.

Если повторить ту же самую процедуру по отношению к группам нежелающих голосовать за кандидата X и затруднившихся с ответом, то получим следующие данных (табл. 21).

Т а б л и ц а 21

**Показатели распределения признака
«Намерение голосовать за кандидата X» после ремонта выборки**

Намерение голосовать за кандидата X	% (мужчины)	Абсолютные показатели (чел.)	% (женщины)	Абсолютные показатели (чел.)	Всего (чел.)
Да	50	225	70	385	610 (61 %)
Нет	35	158	20	110	268 (26,8 %)
Не знаю	15	67	10	55	122 (12,2 %)
Всего	100 %	450	100 %	550	1000

Как видно, после ремонта результаты изменились. Доля собирающихся голосовать за кандидата X уменьшилась на 3 % (с 64 до 61 %). Доли людей, имеющих другие намерения, увеличились. Это связано с уменьшением количества женщин в отремонтированном массиве. Женщины же чаще, чем мужчины, собирались голосовать за данного кандидата.

Ремонт выборки более правильно осуществлять по нескольким сопряженным признакам. Так, при двух признаках (например, пол и возраст) выравнивание будет осуществляться относительно таких значений двухмерного признака, как мужчины до 30 лет,

мужчины 30–49 лет, женщины 40–49 лет и т. д. Естественно, для этого надо иметь информацию о численности данных групп в генеральной совокупности. При таком методе уравниваются не только пропорции по каждому признаку в отдельности (полу и возрасту), но и по каждому из сочетаний значений этих признаков.

Дефицитным единицам *не должны* присваиваться очень большие весовые коэффициенты, особенно при относительно небольшом объеме выборки.

Так, например, если при 400 опрошенных доля лиц с общим средним образованием составила 2 % (8 человек), а в генеральной совокупности этот показатель составил 20 %, то, присвоив данному значению признака весовой коэффициент «10», мы получим группу из 80 респондентов с общим средним образованием. При этом информация о предпочтениях всего лишь восьми индивидов будет распространена на группу из восьмидесяти человек, которая составляет пятую часть от объема выборочной совокупности.

Перевзвешивание данных осуществляется не только по причине обнаружившихся смещений в выборке. Потребность в данной процедуре может возникнуть и при непропорциональном стратифицированном отборе. Когда объем выборки из страты небольшой численности увеличивается в целях ее сравнения с другими стратами, в любом случае нужно получить информацию о генеральной совокупности в целом. Поэтому целесообразно использовать для сравнительного анализа страт непропорциональную выборку, а для получения оценок параметров генеральной совокупности – отремонтированную, в которой восстановлены правильные пропорции страт.

Нужно отметить, что цель социолога состоит в достижении *минимальной ошибки* выборки, относящейся к *исследуемым переменным*. Это переменные, измерение которых входит в задачи исследования. Например, это мнения, оценки, действия обследуемых, относящиеся к предмету исследования. Распространенность данных мнений и оценок в *генеральной совокупности* не может быть зафиксирована социологом, именно поэтому проводится выборочное обследование.

Что касается другой группы переменных, то они служат *средствами* минимизации ошибки по исследуемым переменным, то есть выполнения основной цели выборочных процедур. Информация о распределении данных переменных может быть получена социологом, как уже указывалось выше, из заслуживающих доверия источников. Данные признаки должны оказывать существенное влияние на исследуемые переменные.

Нужно учитывать тот факт, что ремонт выборки не дает полной гарантии устранения систематических смещений по *исследуемым переменным*.

Вопросы и задания для самоконтроля

1. От чего зависит размер предельной ошибки случайной выборки?
2. Как осуществляется апостериорный контроль ошибок репрезентативности?
3. С какой целью осуществляется ремонт выборки?
4. Как можно определить величину систематического смещения, вызванного наличием недоступных и труднодоступных единиц?
5. Перечислите источники систематических ошибок случайных видов отбора.
6. Перечислите источники систематических ошибок квотной выборки.
7. Приведите примеры появления систематической ошибки, вызванной действиями анкетера.
8. Приведите примеры появления систематической ошибки, вызванной действиями социолога.

ПОСЛЕСЛОВИЕ

В заключение хотелось бы отметить, что выборочный метод – это не только учебная дисциплина, но и сфера научных изысканий. Хотелось бы остановиться на тех проблемах, которые требуют дополнительного изучения в рамках методических исследований в данной сфере.

Методы формирования выборки являются средствами получения адекватной информации о генеральной совокупности и оценки ее параметров. Поэтому особенности отбора единиц оказывают влияние на такой этап социологического исследования, как статистический анализ данных. Применение любых самых сложных методов анализа данных окажется совершенно бесполезным, а иногда даже вредным, если выборочные процедуры были осуществлены некачественно. Поэтому требуется дополнительное изучение взаимосвязи используемых выборочных процедур и применяемых математических методов анализа данных.

Нужно отметить, что необходимо изучать особые проблемы формирования выборки в отдельных сферах человеческой деятельности – производственной, электоральной активности, деятельности в учреждениях высшего образования, торговых учреждениях.

Также есть необходимость *анализа методического опыта* проведения *репрезентативных* исследований на основе уличных опросов. Уличное экспресс-интервью может быть использовано как *ограниченная рядом жестких условий* процедура, не допускающая произвола анкетера и осуществляемая в рамках реализации двухступенчатой общегородской выборки.

Требуют дополнительных методических исследований факторы формирования *систематических ошибок* выборки в разных видах исследований.

Следует обратить внимание на наиболее частые ошибки, связанные с формированием выборочной совокупности в учебных исследованиях:

1. Объединение в одной выборочной совокупности групп, принадлежащих к разным генеральным совокупностям. При этом формулируются статистические выводы, относящиеся к выборке в целом, что является неправомерным.

2. Полностью не учитываются финансовые и организационные возможности исследователя при принятии решения о типе выборочной процедуры.

3. Недостаточно используются статистические данные о структуре генеральной совокупности с целью расчета апостериорной ошибки выборки и предотвращения систематических смещений.

4. Необходимо обратить внимание на недостаточный учет требований репрезентативности выборки в процессе проведения сравнительных исследований.

ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Тестовые вопросы и задания

А. Укажите тип выборочной процедуры. На основе данных государственной статистики рассчитана численность всех половозрастных групп в выборке в соответствии с пропорциями в генеральной совокупности. Анкетерам выданы задания на опрос респондентов определенного пола, возраста и образования. Эта выборка является:

1. Простой случайной.
2. Систематической.
3. Гнездовой.
4. Стратифицированной.
5. Квотной.

Б. При формировании общегородской выборки по таблице случайных чисел отобрано 40 избирательных участков. На каждом из участков из алфавитных списков с определенным шагом отобран каждый второй избиратель. Эта выборка является:

1. Одноступенчатой.
2. Двухступенчатой.
3. Трехступенчатой.

В. Генеральная совокупность – население города с численностью населения, входящего в объект исследования (население в возрасте 15 лет и старше), составило 400 тысяч человек. Предмет исследования – отношение населения к теневой экономике. Время опроса составляет 10 минут. Объем выборки – 500 человек. Обеспеченность финансовыми ресурсами высокая.

Какой тип выборочной процедуры целесообразнее всего предпочесть?

Какие действия должен предпринять социолог при формировании данной выборки?

Г. Генеральная совокупность – совокупность учеников 11-х классов одного из административных районов города. Объем генеральной сово-

купности – 10 000 человек. Обеспеченность финансовыми ресурсами невысокая.

Какой тип выборочной процедуры целесообразнее всего предпочесть?

Какие действия должен предпринять социолог при формировании данной выборки?

Д. Обведите номера суждений, являющихся верными.

1. Отбор первого встречного на улице – это разновидность случайного отбора.

2. Чем больше гнезд отбирается, тем ошибка гнездовой выборки выше.

3. Если город разделить на районы и из каждого района по таблице случайных чисел отобрать респондентов, то выборка двухступенчатая.

4. Связь между исследуемыми признаками и признаками расслоения должна быть сильной.

5. Для реализации стратифицированной выборки нужно иметь список всех элементов, входящих в каждую страту.

6. Между квотными группами должны быть сильные различия по исследуемым признакам.

7. Чем больше ступеней в выборке, тем ошибка репрезентативности меньше.

8. В гнездах должен осуществляться сплошной опрос респондентов.

9. Рассчитать предельную ошибку квотной выборки можно по специальной формуле.

10. Чем больше гнезда похожат друг на друга, тем ошибка гнездовой выборки меньше.

11. Отбор из списка элементов генеральной совокупности каждого десятого респондента – систематический отбор.

12. На первой ступени случайной многоступенчатой выборки можно отобрать 5–6 единиц.

13. Для построения квотной выборки нужно иметь основу выборки.

14. Если генеральной совокупностью являются женщины-рабочие в возрасте 30–50 лет, то выборка квотная.

15. Чем однороднее каждая из страт по исследуемым признакам, тем ошибка выборки меньше.

16. Случайная выборка требует повторных посещений отсутствующих респондентов.

17. Таблицы Киша применяются при случайном отборе респондента в семье.

18. Квотная выборка — разновидность вероятностного отбора.

19. Если единицы отбора совпадают с единицами наблюдения, то отбор одноступенчатый.

20. Ремонт выборки по социально-демографическим признакам гарантирует минимизацию ошибок по исследуемым признакам.

21. Если в выборке обнаружено избыточное количество анкет одной из социально-демографических групп, то коэффициент перевзвешивания для этой группы должен составлять величину, превышающую единицу.

22. Систематическую ошибку можно уменьшить, увеличив число опрошенных.

ПРАКТИЧЕСКИЕ ЗАДАНИЯ

1. Сформируйте таблицу квотной выборки. Объем выборки – 700 человек. Объем генеральной совокупности – 10 000 человек.

Возраст	Высшее образование	Среднее специальное образование	Среднее общее образование
18–29 лет	800	600	400
30–49 лет	1200	800	900
50–59 лет	1000	400	1900
60 и старше	500	900	600
Всего	3500	2700	3800

2. Осуществлена простая случайная выборка объемом 300 человек. Объем генеральной совокупности – 200 000 человек. По результатам опроса, полностью удовлетворенные работой составили 47 % выборочной совокупности. Рассчитайте предельную ошибку выборки с доверительной вероятностью $P = 0,954$.

3. Рассчитайте коэффициенты перевзвешивания. Определите численность образовательных групп после ремонта. Сколько процентов составят доверяющие руководству в выборочной совокупности до и после ремонта? Объем выборочной совокупности – 1100.

Образование	Генеральная совокупность (в процентах)	Выборочная совокупность (в процентах)	Процент доверяющих руководству предприятия в группе
Высшее и незаконченное высшее	30	15	60
Среднее специальное	20	25	40
Среднее общее	50	60	15

4. Сформируйте таблицу квотной выборки. Объем выборки – 600 человек. Объем генеральной совокупности – 20 000 человек.

Возраст	Рабочие 1–2 разряда	Рабочие 3–4 разряда	Рабочие 5–6 разряда
18–29 лет	1600	1200	800
30–49 лет	2400	1600	1800
50–59 лет	2000	800	3800
60 и старше	1000	1800	1200
Всего	7000	5400	7600

5. Осуществлена простая случайная выборка объемом 500 человек. Объем генеральной совокупности – 200 000 человек. По результатам опроса, поддерживающие президента составили 75 % выборочной совокупности. Рассчитайте предельную ошибку выборки с доверительной вероятностью $P = 0,954$.

6. Рассчитайте коэффициенты перевзвешивания. Определить численность групп после ремонта $n = 1400$.

Образование	Генеральная совокупность (в процентах)	Выборочная совокупность (в процентах)
Высшее и незаконченное высшее	10	20
Среднее специальное	45	40
Среднее общее	45	40

7. Отберите из списка районов области 10 районов с вероятностью, пропорциональной численности населения в районе.

Суммарная численность населения – 722 701 человек.

Районы	Численность населения в возрасте 18 лет и старше
1. Кузнецкий	40893
2. Колышлейский	26421

Продолжение табл.

Районы	Численность населения в возрасте 18 лет и старше
3. Камешкирский	12881
4. Лопатинский	14911
5. Лунинский	19477
6. Малосердобинский	10451
7. Мокшанский	28735
8. Наровчатский	12013
9. Неверкинский	17372
10. Нижнеломский	43509
11. Никольский	36290
12. Пачелмский	16401
13. Пензенский	51013
14. Сердобский	55499
15. Сосновоборский	17724
16. Спасский	12574
17. Тамалинский	17190
18. Шемышейский	17588
19. Бежаницкий	14793
20. Великолукский	22690
21. Гдовский	15810
22. Новобурасский	16717
23. Новоузенский	33264
24. Озинский	22403
25. Перелобский	17798
26. Петровский	14764

О к о н ч а н и е т а б л.

Районы	Численность населения в возрасте 18 лет и старше
27. Питерский	18361
28. Пугачевский	20449
29. Ровенский	17640
30. Романовский	16822
31. Ртищевский	17648
32. Самойловский	22600

8. Генеральная совокупность – население города. На основе информации о составе генеральной совокупности районов (см. табл. 1–7) составьте таблицу квотной выборки. Объем выборки – 1000 человек. Квотный признак – трехмерная переменная «Половозрастная принадлежность, сопряженная с районом проживания».

Т а б л и ц а 1

Орджоникидзевский район	Мужчины	Женщины
15–19	7100	6500
20–24	11800	11600
25–29	12200	14100
30–34	11600	13400
35–39	11200	12100
40–44	8800	10700
45–49	10400	11920
50–54	9600	13600
55–59	8700	13300

Т а б л и ц а 2

Чкаловский район	Мужчины	Женщины
15–19	5500	5100
20–24	8240	8310
25–29	9710	10820
30–34	11600	12400
35–39	10800	11950
40–44	8870	10400
45–49	10100	11900
50–54	8970	13100
55–59	8880	11950

Т а б л и ц а 3

Октябрьский район	Мужчины	Женщины
15–19	1950	1824
20–24	5536	5490
25–29	5750	6500
30–34	5500	5850
35–39	4540	5590
40–44	4200	5010
45–49	4236	5230
50–54	4240	5850
55–59	4210	6100

Т а б л и ц а 4

Ленинский район	Мужчины	Женщины
15–19	3650	3510
20–24	5902	6015
25–29	7150	8507
30–34	6403	8050
35–39	6102	7560
40–44	5402	6400
45–49	5300	6800
50–54	5250	7203
55–59	4900	7120

Т а б л и ц а 5

Кировский район	Мужчины	Женщины
15–19	4913	4900
20–24	4670	8800
25–29	9950	11400
30–34	8750	9910
35–39	7870	9750
40–44	7020	7400
45–49	7250	8700
50–54	7400	10410
55–59	10050	4250

Т а б л и ц а 6

Верх-Исетский район	Мужчины	Женщины
15–19	4720	4520
20–24	7230	7300
25–29	9041	10440
30–34	7890	10141
35–39	7200	9041
40–44	6700	7600
45–49	6700	8400
50–54	6850	9460
55–59	6280	9050

Т а б л и ц а 7

Железнодорожный район	Мужчины	Женщины
15–19	3175	3100
20–24	5200	5400
25–29	6350	7050
30–34	5550	6050
35–39	5400	5950
40–44	4300	4750
45–49	4500	5270
50–54	4858	6350
55–59	4250	6335

9. Рассчитайте предельную ошибку доли учеников, интересующихся политикой (выборка гнездовая). Объем генеральной совокупности – 400 классов.

Классы	Число учеников в классе	Число интересующихся политикой в классе	Процент интересующихся политикой в классе
1	22	4	18,2
2	20	7	35
3	31	6	19,4
4	29	10	34,5
5	25	6	24
6	30	6	20
7	22	3	13,6
8	32	7	21,9
9	24	5	20,8
10	26	7	26,9
11	28	6	21,4
12	18	4	22,2
13	29	10	34,5
14	23	5	21,7
15	30	5	16,7
16	26	7	26,9
17	27	6	22,2
18	23	7	30,4
19	34	5	14,7
20	32	3	9,4
21	28	11	39,3

О к о н ч а н и е т а б л.

Классы	Число учеников в классе	Число интересующихся политикой в классе	Процент интересующихся политикой в классе
22	36	5	13,9
23	23	4	17,4
24	31	6	19,4
25	29	7	24,1
26	25	8	32
27	28	7	25
28	26	6	23,1
29	29	9	31,0
30	24	8	33,3
31	28	4	14,3
32	19	2	10,5
33	26	11	42,3
34	34	6	17,6
35	15	3	20
Всего	932	216	18,2

10. Генеральная совокупность – совокупность работников промышленных предприятий административного района города, численность которых составляет 10 000 человек. Признак расслоения – должность работника.

Должностные группы работников:

- рабочие;
- инженерно-технические работники;
- младший обслуживающий персонал.

Сколько процентов в выборке составляют работники, желающие уволиться с предприятия? Необходимо рассчитать предельную ошибку

стратифицированной выборки относительно доли работников, высказавших данное желание.

При расчетах используйте данную таблицу.

Страты	Число единиц в страте (генеральная совокупность)	Число единиц в страте (выборочная совокупность)	Число людей, желающих уволиться с предприятия, в страте
Рабочие	5000	250	60
Инженерно- технические работники	4000	200	14
Младший обслуживающий персонал	1000	50	10
Всего	10000	500	84

11. На избирательном участке – пять домов многоквартирной застройки, расположенных на одной улице. Рассчитайте общее количество квартир в домах, укажите накопленные частоты. Определите шаг отбора. Определите номер первой обследуемой квартиры по таблице случайных чисел. Отберите для проведения опроса 10 квартир при помощи систематического отбора, используя таблицу.

Номер дома	Номера квартир в доме
53	1–347
55	1–282
57	1–445
59	1–275
61	1–251

12. Из списка избирательных участков административного района города отберите 11 избирательных участков при помощи простого случайного отбора.

Номер избирательного участка	Адреса
1570	Улицы: Луначарского, 15, 17, 21, 21а, 33; Короленко, 5, 9; Восточная, 3, 5, 6, 8, 10, 12, 14, 16, 22. Переулок Трамвайный, 2, 2а, 8
1571	Улицы: Луначарского, 22/13, 34, 36, 38, 40; Мамина-Сибиряка, 8, 10, 23, 25, 45; Азина, 15, 21, 23
1572	Улицы: Луначарского, 48, 49, 50, 51, 53, 53а, 55, 57, 60; Короленко, 4, 8, 8а, 10, 10а; Восточная, 20, 24, 26а, 28
1573	Улицы: Шевченко, 19, 21, 23, 25, 25а, 27, 29, 29а, 31, 33, 35; Короленко, 14; Мамина-Сибиряка, 57а
1574	Улицы: Свердлова, 2, 4, 6, 10, 14, 22, 30, 34, 56, 56а, 58; Шевченко, 9; Лермонтова, 4; Шевченко, 9ж, 11, 11/42, 15; Мамина-Сибиряка, 40, 51, 59, 71
1575	Улицы: Мамина-Сибиряка, 2, 2а; Азина, 18, 18а, 20/1, 20/2, 20/3, 20/4, 26, 30, 39; Свердлова, 60, 62, 66; Челюскинцев, 110, 110а
1576	Улицы: Еремина, 15; Мельковская, 3, 9, 11, 13; Лермонтова, 15, 17, 17а, 19, 21, 25, 27, 28; Физкультурников, 20, 27, 30, 37; Испанских рабочих, 8. Переулок Красный, 13, 15, 17, 19
1577	Улицы: Мельковская, 26; Братьев Быковых, 5/29, 7, 32, 34; Еремина, 3; Испанских Рабочих, 31/26, 35, 45; Челюскинцев, 70, 88; Азина, 46, 53, 55, 59
1578	Улица Челюскинцев, 19, 21, 23, 25, 27, 29, 31, 33, 33а. Переулок Невьянский, 1
1579	Улицы: Испанских Рабочих, 26, 28, 40, 40/53; Челюскинцев, 92; Свердлова, 11, 15, 25, 27; Азина, 40, 42, 45, 47; Братьев Быковых, 18

Продолжение табл.

Номер избирательного участка	Адреса
1580	Улицы: Некрасова, 2, 6, 8, 12, 14, 16; Челюскинцев, 9, 11; Печерская, 2, 3, 4, 4а, 6, 6а; Челюскинцев, 11/7
1581	Улицы: Гражданская, 2, 2а, 4, 17; Кимовская, 4, 6, 8, 10; Машинистов, 2, 3, 10, 12, 12а, 14, 19а
1582	Улицы: Колмогорова, 54а, 56, 58, 58/2д, 60, 62, 64, 68; Машинистов, 4, 4а; Одинарка, 1, 2е, 3
1583	Улицы: Тюменская, 18, 20; Завокзальная, 2, 3, 4, 4а; Тагильская, 15, 15а
1584	Улица Ольховская, 23, 25/1, 25/2, 27/1, 27/2
1585	Улицы: Бебеля, 156, 158; Пехотинцев, 10
1586	Улица Бебеля, 146, 148, 152, 154
1587	Улицы: С. Перовской, 119; Пехотинцев, 12, 18
1588	Улица С. Перовской, 113, 115, 117
1589	Улицы: Пехотинцев, 13, 17, 19, 21, 21б; Автомагистральная, 15, 19, 23
1590	Улицы: Автомагистральная, 21, 25, 27, 29, 31, 33, 35; Пехотинцев, 9, 11; Ольховская, 2/1, 2/2; Уральских коммунаров 2, 2а, 5, 9, 15, 17, 19
1591	Улицы: Пехотинцев, 5, 7; Бебеля с 166 по 172, 172а, 173, 174, 176, 176а, 182а. Теплоходный проезд, 5, 7, 9, 11
1592	Улицы: Бебеля, с 117 по 127, 162, 164; Лесная, 1, 1а, 4а, 5, 5а, 17, 38, 38а, 39, 40; Таватуйская, 4, 12/1
1593	Улицы: Техническая, 26, 27, 28, 31, 31а, 33, 35, 35а; Минометчиков, 5, 7, 26, 27, 28, 30; Таватуйская, 6

О к о н ч а н и е т а б л .

Номер избирательного участка	Адреса
1594	Улицы: Техническая, 12, 16, 18, 20, 22/2; Таватуйская, 1а, 1б, 1в; Седова, 17
1595	Улицы: Надеждинская, 3, 8, 10, 12а, 12б, 12в, 12г, 14, 20; Седова, 23, 25
1596	Улицы: Техническая, 22/1, 22/324, 41, 43, 45, 47, 47а, 49; Минометчиков, 34, 36, 38; Седова, 26; Таватуйская, 2
1597	Улицы: Техническая, 36, 38, 38а, 40, 42, 42а; Седова, 30, 31, 38, 38а, 47, 57; Надеждинская, 9, 11, 13; Сортировочная, 9, 13
1598	Улицы: Седова, 42, 44, 44а; Надеждинская, 17, 19, 21, 23
1599	Улицы: Таватуйская, 1, 5; Надеждинская, 17, 19, 21, 23, 25; Коуровская, 12, 16; Седова 46,48
1600	Улицы: Техническая, 44, 44а, 46, 48, 50, 52, 53, 56, 58, 60; Седова, 33, 37, 39, 48а, 52, 14, 18, 21, 23
1601	Улицы: Сортировочная, 1, 4; Минометчиков; Седова, 43, 45, 56, 61, 48; Коуровская, 6, 14, 17, 22, 24, 26, 28; Таватуйская, 9, 19; Техническая, 51, 54, 55, 65, 67; Минометчиков, 40а, 42, 44
1602	Улицы: Коуровская, 11; Минометчиков; Таежная, 2, 48; Ватугина, 8; Техническая, 58, с 62 по 70; Минометчиков, 27, 56, 58, 62; Соликамская, 7

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

О с н о в н а я

Васильева Э. К. Выборочный метод в социально-экономической статистике : учеб. пособие для студентов вузов / Э. К. Васильева, М. М. Юзбашев. М. : Финансы и статистика: ИНФРА-М, 2010.

Добреньков В. И. Методы социологического исследования : учебник для вузов по специальности 020300 «Социология» / В. И. Добреньков, А. И. Кравченко. М. : ИНФРА-М, 2009.

Тихонова Е. В. Методология и методы социологического исследования : учебник для студентов учреждений высш. проф. образования / Е. В. Тихонова. М. : Академия, 2012.

Д о п о л н и т е л ь н а я

Девятко И. Ф. Методы социологического исследования : учеб. пособие / И. Ф. Девятко. М. : Университет, 2002.

Елисеева И. И. Общая теория статистики : учебник для вузов по направлению и специальности «Статистика» / И. И. Елисеева, М. М. Юзбашев ; под ред. И. И. Елисеевой. М. : Финансы и статистика, 2006.

Ефимова М. Р. Практикум по общей теории статистики : учеб. пособие для вузов / М. Р. Ефимова, О. И. Ганченко, Е. В. Петрова. М. : Финансы и статистика, 2007. 367 с.

Йейтс Ф. Выборочный метод в переписях и обследованиях. М. : Статистика, 1965.

Кокрен У. Методы выборочного исследования / У. Кокрен ; пер. с англ. И. М. Сониной ; под ред. А. Г. Волкова ; предисл. к рус. пер. Н. К. Дружинина. М. : Статистика, 1976.

Михок Г. Выборочный метод и статистическое оценивание / Г. Михок, В. Урсяну ; пер. с румын. В. М. Остиану ; под ред. В. Ф. Матвеева. М. : Финансы и статистика, 1982. 245 с.

Нозль Э. Массовые опросы: Введение в методику демоскопии / пер. с нем. ; общ. ред. и вступ. ст. Н.С.Мансурова. М. : Прогресс, 1978.

Паниотто В. И. Качество социологической информации: (методы оценки и процедуры обеспечения) / В. И. Паниотто ; отв. ред. В. Е. Хмелько. Киев : Наук. думка, 1986.

Проектирование и организация выборочного социологического исследования. М. : Ин-т социол. исследований, 1977.

Рогозин Д. М. Результативность телефонного опроса в зависимости от ограничений на выбор респондента внутри домохозяйства / Д. М. Рогозин // Социол. журнал. 2005. № 3. С. 52–85.

Сотникова Г. Н. Методы контроля и компенсации смещений от недоступных единиц в выборочном социологическом исследовании / Г. Н. Сотникова // Применение математических методов и ЭВМ в социологических исследованиях. М. : ИСИ АН СССР, 1982. С. 102–117.

Территориальная выборка в социологических исследованиях / И. Б. Мучник, Е. С. Петренко, Е. Э. Сеницын, Т. М. Ярошенко ; отв. ред. Т. В. Рябушкин. М. : Наука, 1980.

Тулебаева А. А. Репрезентативность выборки в исследовании социальных объектов : дисс. ... канд. социол. наук / А. А. Тулебаева. Екатеринбург, 2010.

Черепанов Е. В. Стохастическое описание выборочного метода / Е. В. Черепанов // Социология 4М. 2007. № 25. С. 167–189.

Чурилов Н. Н. Проектирование выборочного социологического исследования (некоторые методологические и методические проблемы) / Н. Н. Чурилов ; АН УССР, Ин-т философии ; отв. ред. В. И. Паниотто. Киев : Наукова думка, 1986.

Шварц Г. Выборочный метод : рук. по применению стат. методов оценивания / Г. Шварц ; пер. с нем. Я. Ш. Паппэ ; под ред. И. Г. Венецкого, В. М. Ивановой. М. : Статистика, 1978.

Шереги Ф. Э. Применение метода квот в выборочном социологическом исследовании / Ф. Э. Шереги // СОЦИС. 1975. № 3. С. 74–85.

Шляпентох В. Э. Проблемы репрезентативности социологической информации: случайные и неслучайные выборки в социологии / В. Э. Шляпентох. М. : Статистика, 1976.

Шляпентох В. Э. Проблемы качества социологической информации: достоверность, репрезентативность, прогностический потенциал / В. Э. Шляпентох. М. : Центр социального прогнозирования, 2006.

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Т а б л и ц а 1

Таблицы случайных чисел¹²

10097	32533	76520	13586	34673	54876	80959	09117	39292
37542	04805	64894	74296	24805	24037	20636	10402	00822
08422	68953	19645	09303	23209	02560	15953	34764	35080
99019	02529	09376	70715	38311	31165	88676	74397	04436
12807	99970	80157	36147	64032	36653	98951	16877	12171
66065	74717	34072	76850	36697	36170	65813	39885	11199
31060	10805	45571	82406	35303	42614	86799	07439	23403
85269	77602	02051	65692	68665	74818	73053	85247	18623
63573	32135	05325	47048	90553	57548	28468	28709	83491
73796	45753	03529	64778	35808	34282	60935	20344	35273
98520	17767	14905	68607	22109	40558	60970	93433	50500
11805	05431	39808	27732	50725	68248	29405	24201	52775
83452	99634	06288	98083	13746	70078	18475	40610	68711
88685	40200	86507	58401	36766	67951	90364	76493	29609
99594	67348	87517	64969	91826	08928	93785	61368	23478
65481	17674	17468	50950	58047	76974	73039	57186	40218

¹² Источник: The RAND Corporation. A Million Random Digits with 100 000 Normal Deviates. N. Y. : Free Press, 1966. P. 1.

О к о н ч а н и е т а б л . 1

80124	35635	17727	08015	45318	22374	21115	78253	14385
74350	99817	77402	77214	43236	00210	45521	64237	96286
69916	26803	66252	29148	36936	87203	76621	13990	94400
09893	20505	14225	68514	46427	56788	96297	78822	54382
91499	14523	68479	27686	46162	83554	94750	89923	37089
80336	94598	26940	36858	70297	34135	53140	33340	42050
44104	81949	85157	47954	32979	26575	57600	40881	22222
12550	73742	11100	02040	12860	74697	96644	89439	28707
63606	49329	16505	34484	40219	52563	43651	77082	07207
61196	90446	26457	47774	51924	33729	65394	59593	42582
15474	45266	95270	79953	59367	83848	82396	10118	33211

Т а б л и ц а 2

**Значение интеграла вероятностей
нормального закона распределения**

Сотые доли										
	0	1	2	3	4	5	6	7	8	9
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3752
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4331	4448
0,6	4515	4581	4647	4713	4778	4843	4909	4971	5035	5098
0,7	5161	5223	5285	5346	5406	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	6817	6875	6923	6970	7017	7063	7109	7154	7199	7243

Окончание табл. 2

Сотые доли										
	0	1	2	3	4	5	6	7	8	9
1,1	7699	7330	7373	7415	7547	7499	7540	7580	7620	7660
1,2	8064	7737	7775	7813	7850	7887	7923	7959	7995	8030
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8788	8812	8936	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9108	9127	9146	9164	9182	9199	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9425	9438	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9608	9615	9625	9634
2,1	9643	9652	9660	9669	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9755	9762	9768	9774	9780
2,3	9785	9791	9797	9802	9807	9812	9817	9822	9827	9932
2,4	9836	9840	9845	9849	9853	9857	9861	9866	9869	9872
2,5	9876	9879	9883	9886	9889	9898	9895	9898	9901	9934
2,6	9907	9909	9912	9915	9917	9920	9922	9924	9926	9929
2,7	9931	9933	9935	9937	9939	9940	9942	9944	9946	9947
2,8	9949	9950	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	9973	9973	9974	9975	9976	9977	9977	9978	9979	9980

Т а б л и ц а 3

Критические точки распределения Стьюдента

Число степеней свободы df	Доверительная вероятность P	
	0,95	0,99
1	12,706	63,657
2	4,3027	9,9248
3	3,1825	5,8409
4	2,7764	4,6041
5	2,5706	4,0321
6	2,4469	3,7074
7	2,3646	3,4995
8	2,3060	3,3554
9	2,2622	3,2498
10	2,2281	3,1693
11	2,2010	3,1058
12	2,1788	3,0545
13	2,1604	3,0123
14	2,1448	2,9768
15	2,1315	2,9467
16	2,1199	2,9208
17	2,1098	2,8982
18	2,1009	2,8784
19	2,0930	2,8609
20	2,0860	2,8453
21	2,0790	2,8310
22	2,0739	2,8188

Продолжение табл. 3

Число степеней свободы df	Доверительная вероятность P	
	0,95	0,99
23	2,0687	2,8073
24	2,0639	2,7969
25	2,0595	2,7874
26	2,059	2,778
27	2,0518	2,7707
28	2,0484	2,7633
29	2,0452	2,7564
30	2,0423	2,7500
32	2,0360	2,7380
34	2,0322	2,7284
36	2,0281	2,7195
38	2,0244	2,7116
40	2,0211	2,7045
42	2,018	2,6980
44	2,0154	2,6923
46	2,0129	2,6870
48	2,0106	2,6822
50	2,0086	2,6778
55	2,0040	2,6680
60	2,0003	2,6603
65	1,997	2,6536
70	1,9944	2,6479
80	1,9900	2,6380

О к о н ч а н и е т а б л . 3

Число степеней свободы df	Доверительная вероятность P	
	0,95	0,99
90	1,9867	2,6316
100	1,9840	2,6259
120	1,9719	2,6174
150	1,9759	2,6090
200	1,9719	2,6006
250	1,9695	2,5966
300	1,9679	2,5923
400	1,9659	2,5882
500	1,9640	2,7850

ОГЛАВЛЕНИЕ

Предисловие	3
РАЗДЕЛ 1. Выборочное обследование как разновидность не сплошного обследования. Основные понятия теории выборки	5
РАЗДЕЛ 2. Случайные и неслучайные виды отбора	14
РАЗДЕЛ 3. Многоступенчатая выборка	37
РАЗДЕЛ 4. Случайные и систематические ошибки выборки	56
Послесловие	91
Задания для самостоятельной работы	93
Практические задания	96
Список рекомендуемой литературы	109
Приложение	111

Учебное издание

Могильчак Елена Львовна

**ВЫБОРОЧНЫЙ МЕТОД
В ЭМПИРИЧЕСКОМ
СОЦИОЛОГИЧЕСКОМ
ИССЛЕДОВАНИИ**

Учебное пособие

Зав. редакцией *М. А. Овечкина*
Редактор *Е. Е. Крамаревская*
Корректор *Е. Е. Крамаревская*
Компьютерная верстка *Г. Б. Головиной*

План изданий 2015 г. Подписано в печать 17.06.15.
Формат 60×84/16. Бумага офсетная. Гарнитура Times.
Уч.-изд. л. 6,00. Усл. печ. л. 6,98. Тираж 100 экз. Заказ 218.

Издательство Уральского университета
620000, Екатеринбург, ул. Тургенева, 4

Отпечатано в Издательско-полиграфическом центре УрФУ
620000, Екатеринбург, ул. Тургенева, 4
Тел.: + (343) 350-56-64, 350-90-13
Факс +7 (343) 358-93-06
E-mail: press-urfu@mail.ru

Для заметок



МОГИЛЬЧАК ЕЛЕНА ЛЬВОВНА

Кандидат философских наук, доцент кафедры прикладной социологии Уральского федерального университета. Читает курсы «Выборочный метод в социологическом исследовании», «Теория измерений и анализ данных», «Экономическая социология», «Социальная статистика». Является автором научных публикаций в журналах «Социологические исследования», «Социологический журнал», *Russian Education and Society*, «Университетское управление: практика и анализ». Сфера научных интересов – социально-экономические проблемы молодежи, социальное самочувствие и ценностные ориентации личности, культура труда и потребления. Основные исследовательские проекты: «Молодежь: образование и жизненные пути» (1989–1995), «Мониторинг региональных проблем доступности высшего образования» (2005–2007), «Студент и экономика» (2011–2013).